Semi-parametric estimation of the hazard function in a model with covariate measurement error

Marie-Laure Martin-Magniette, Marie-Luce Taupin

SEMI-PARAMETRIC ESTIMATION OF THE HAZARD FUNCTION IN A MODEL WITH COVARIATE MEASUREMENT ERROR

MARIE-LAURE MARTIN-MAGNIETTE[1,2], MARIE-LUCE TAUPIN[3]

[1]I.N.A. Paris-Grignon, Mathématique et Informatique Appliquées, Paris,
[2] I.N.R.A., Unité de Recherche en Genomique Végétale, Evry, France,
[3]Université Paris-Sud, Orsay and Université Paris René Descartes, IUT de Paris

Abstract. We consider a model where the failure hazard function, conditional on a time-independent covariate Z is given by $R(t,\theta^0|Z)=\eta_{\gamma^0}(t)f_{\beta^0}(Z)$, with $\theta^0=(\beta^0,\gamma^0)^\top\in\mathbb{R}^{m+p}$. The baseline hazard function η_{γ^0} and relative risk f_{β^0} belong both to parametric families. The covariate Z is measured with an error through an additive error model $U=Z+\varepsilon$ where ε is a random variable, independent from Z, with known density f_ε. We observe a n-sample (X_i,D_i,U_i), $i=1,\dots,n$, where X_i is the minimum between the failure time and the censoring time, and D_i is the censoring indicator. We aim at estimating θ^0 in presence of the unknown density g of the covariate Z using the observations (X_i,D_i,U_i), $i=1,\dots,n$. Our estimation procedure based on least squares criterion provide two estimators of θ^0. The first one is obtained by minimizing an estimation of the least squares criterion where g is estimated by density deconvolution. We give upper bounds for its risk that depend on the smoothness properties of f_ε and $f_\beta(z)$ as a function of z. We derive from this construction sufficient conditions that ensure the $\sqrt{n}$-consistency. The second estimator is constructed under conditions ensuring that the least squares criterion can be directly estimated with the parametric rate. We propose a deep study of examples considering various type of relative risks f_β and various types of error density f_ε. We show in particular that in the Cox model and in the excess risk model, the estimators are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0 whatever is f_ε.

Abstract. Considérons un modèle à risque instantané modélisé par la relation $R(t,\theta^0|Z)=\eta_{\gamma^0}(t)f_{\beta^0}(Z)$, où $\theta^0=(\beta^0,\gamma^0)^\top\in\mathbb{R}^{m+p}$. Le risque de base η_{γ^0} et la fonction de risque relatif f_{β^0} appartiennent à des familles paramétriques. La covariable Z est mesurée avec une erreur au travers de la relation $U=Z+\varepsilon$, ε étant une variable aléatoire, indépendante de Z, de densité connue f_ε. Nous disposons d'un n-échantillon (X_i,D_i,U_i), $i=1,\dots,n$ où X_i est le minimum entre le temps de survie et le temps de censure et D_i est l'indicateur de censure. Notre but est d'estimer θ^0, en présence la densité inconnue g, de la covariable Z, en utilisant les observations (X_i,D_i,U_i), $i=1,\dots,n$. Notre méthode d'estimation, fondée sur le critère des moindres carrés nous fournit deux estimateurs. Pour le premier, nous établissons des bornes supérieures du risque dépendant des régularités de la densité des erreurs f_ε et de la fonction de risque relatif, comme fonction de z. Nous en déduisons des conditions suffisantes pour atteindre la vitesse paramétrique. Le deuxième estimateur est construit sous des hypothèses assurant que le critère des moindres carrés peut être estimé à la vitesse paramétrique. Au travers d'exemples, nous étudions les propriétés des estimateurs ainsi que les conditions assurant la $\sqrt{n}$-consistance pour des fonctions de risque relatif et des densités d'erreurs variées. En particulier, dans le modèle de Cox et dans le modèle d'excès de risque, les estimateurs construits sont $\sqrt{n}$-consistants et asymptotiquement gaussiens, quelle que soit la loi des erreurs ε.

Key Words and Phrases: Semiparametric estimation, errors-in-variables model, nonparametric estimation, excess risk model, Cox model censoring, survival analysis.
MSC Classifications (2000): Primary 62G05, 62F12,62N01, 62N02; Secondary 62J02.

[1]*

1. Introduction

In a proportional hazard model the hazard function is defined by

$$R(t, \theta^0|Z) = \eta_{\gamma^0}(t) f_{\beta^0}(Z), \tag{1.1}$$

where η_{γ^0} is the baseline hazard function and f_{β^0} is the relative risk, *i.e.* the risk associated with the value of the covariate Z and relative to the risk under standard condition given by $f_{\beta^0}(0) = 1$. In this paper we consider general relative risk f_β with a special interest in $f_\beta(z) = \exp(\beta z)$ and $f_\beta(z) = 1 + \beta z$ which define respectively the Cox model and the model of excess relative risk. The functions η_{γ^0} and f_{β^0} belong both to parametric families and $\theta^0 = (\beta^0, \gamma^0)^\top$ belongs to the interior of a compact set $\Theta = \mathbb{B} \times \Gamma \subset \mathbb{R}^{m+p}$. To ensure that the hazard function is a positive function, we assume that both are positive functions.

We are interested in the estimation of θ^0 when the covariate Z is measured with error. If Z were measured without error, we would consider a cohort of n individuals during a fixed time interval $[0, \tau]$. For each individual, we would observe a triplet (X_i, D_i, Z_i), where $X_i = \min(T_i, C_i)$ is the minimum between the failure time T_i and the censoring time C_i, $D_i = 1\!\!1_{T_i \leq C_i}$ denotes the failure indicator, and Z_i is the value of the covariate. In this paper we consider that the covariate is mismeasured. For example the covariate Z is a stage of a disease, not correctly diagnosed or a dose of ingested pathogenic agent, not correctly evaluated, so that the error range between the unknown dose and the evaluated dose is sizeable. In this context, the available observation for each individual is the triplet $\Delta_i = (X_i, D_i, U_i)$ where U_i is an evaluation of the unobservable covariate Z_i. The random variables U and Z are related by the error model defined by

$$U = Z + \varepsilon, \tag{1.2}$$

where ε is a centered random variable, independent of Z, T, and C. The density of ε is known and denoted by f_ε. Our aim is thus to estimate the parameter $\theta^0 = (\beta^0, \gamma^0)^\top$ from the n-sample $(\Delta_1, \dots, \Delta_n)$ in the presence of the unknown density g of the unobservable covariate Z, seen as a nuisance parameter belonging to a functional space.

1.1. **Previous known results and ideas.**

Models with measurement errors are deeply studied since the 50's with the first papers of Kiefer and Wolfowitz (1956) and Reiersøl (1950) for regression models with errors-in-variables. We refer to Fuller (1987) and Carroll *et al.* (1995) for a presentation of such models and results related to measurement error models. The interest for survival models when covariates are subject to measurement errors is more recent.

To take into account that the covariate Z is measured with error, the first idea is simply to replace Z with the observation U in the score function defined by

$$L_n^{(1)}(\beta, Z^{(n)}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\frac{f_\beta^{(1)}(Z_i)}{f_\beta(Z_i)} - \frac{\sum_{j=1}^n Y_j(t) f_\beta^{(1)}(Z_j)}{\sum_{j=1}^n Y_j(t) f_\beta(Z_j)} \right) dN_i(t), \tag{1.3}$$

where $N_i(t) = 1\!\!1_{X_i \leq t, D_i = 1}$, $Y_i(t) = 1\!\!1_{X_i \geq t}$, $Z^{(n)} = (Z_1, \dots, Z_n)$, and where $f_\beta^{(1)}$ is the first derivative of f_β with respect to β. We refer to Gill and Andersen (1982) for futher details on (1.3). This method, named the naive method, is known to provide, even in the Cox model, a biased estimator of β^0. This comes from the fact that

$$\lim_{n\to\infty} \mathbb{E}[L_n^{(1)}(\beta^0, U^{(n)})] \neq \lim_{n\to\infty} \mathbb{E}[L_n^{(1)}(\beta^0, Z^{(n)})] = 0.$$

To our knowledge, all previously known results about consistency for the semi-parametric estimation of the hazard function when the covariate is mismeasured are obtained in the Cox model. Let us present those results. Various authors propose estimation procedures based on corrections of the score function $L_n^{(1)}(\beta, U^{(n)})$. Among them, one can cite Kong (1999) who calculates the asymptotic bias of the naive estimator obtained by minimization of $L_n^{(1)}(\beta^0, U^{(n)})$, and defines an adjusted estimator. His estimator is not consistent, but a simulation study indicates that it is less biased than the naive estimator. In the same context, Buzas (1998) proposes an unbiased score function, and shows throughout a simulation study that his method yields to an estimator with a small bias. Following the approach developped first by Stefanski (1989) and Nakamura (1990) for generalized linear models, Nakamura (1992) constructs an approximately corrected partial score likelihood, defined by $L_n^{(1)}(\beta, U^{(n)}) + \sigma^2\beta N(\tau)$, where $N(\tau)$ is the number of failures in the interval $[0,\tau]$ and where ε is a centered Gaussian random variable with variance σ^2. Under the error model defined in (1.2), this correction is based on the facts :

$$\lim_{n\to\infty} \mathbb{E}[L_n^{(1)}(\beta, Z^{(n)})] \text{ only depends on } \mathbb{E}(Z) \text{ and } \mathbb{E}[\exp(\beta Z)], \tag{1.4}$$

$$\mathbb{E}(Z) = \mathbb{E}(U) \tag{1.5}$$

$$\mathbb{E}[\exp(\beta U)] = \mathbb{E}[\exp(\beta Z)]\mathbb{E}[\exp(\beta\varepsilon)]. \tag{1.6}$$

Kong and Gu (1999) prove that the Nakamura (1992)'s estimator is a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of β^0. One can also cite Augustin (2004) who proposes an exact correction of the log-likelihood function.

Again in the Cox model, an extension of the previously mentionned works is presented in Hu and Lin (2002). They obtain a broad class of consistent estimators for the regression parameter when U is measured on all study individuals and the true covariate is ascertained on a randomly selected validation set. A nonparametric correction approach of the partial score function is also developped by Huang and Wang (2000) when repetitions are available.

We point out that those results strongly depend on the exponential form of the relative risk of the Cox model, through the use of (1.4)-(1.6) and the extension of such methods to other relative risks is not concluding. For instance, in the model of excess relative risk without errors, the hazard function is defined by $R(t,\theta^0|Z) = \eta_{\gamma^0}(t)(1+\beta^0 Z)$ and the score function is given by

$$L_n^{(1)}(\beta, Z^{(n)}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left(\frac{Z_i}{1+\beta Z_i} - \frac{\sum_{i=j}^{n} Y_j(t)Z_j}{\sum_{j=1}^{n} Y_j(t)(1+\beta Z_j)}\right) dN_i(t).$$

In this model, the naive method also provides biased estimator of β^0, since

$$\lim_{n\to\infty} \mathbb{E}[L_n^{(1)}(\beta^0, U^{(n)})] \neq \lim_{n\to\infty} \mathbb{E}[L_n^{(1)}(\beta^0, Z^{(n)})] = 0.$$

Indeed, easy calculations combined with the Law of Large Numbers give that the limit $\lim_{n\to\infty} \mathbb{E}[L_n^{(1)}(\beta, Z^{(n)})]$ depends on $\mathbb{E}[Z/(1+\beta Z)]$ whereas the limit $\lim_{n\to\infty} \mathbb{E}[L_n^{(1)}(\beta, U^{(n)})]$ depends on $\mathbb{E}[U/(1+\beta U)]$. Since the error model (1.2) does not provide any expression of $\mathbb{E}[Z/(1+\beta Z)]$ related to $\mathbb{E}[U/(1+\beta U)]$, corrections analogous to the ones proposed in the Cox model cannot be exhibited. In other words, it seems impossible to find a function $\Psi(\beta, U)$, independent of the unknown density g satisfying that $\mathbb{E}(\Psi(\beta,U)) = \mathbb{E}[Z/(1+\beta Z)]$. Consequently the methods proposed in the Cox model, by Nakamura (1992), Kong and Gu (1999), Buzas (1998), Lin (2002), Huang and Wang (2000) or by Augustin (2004) cannot be applied to the model of excess relative risk and *a fortiori* to a model with a general relative risk. As a conclusion, methods based on a correction of the partial score likelihood (1.3) where Z is replaced with U seem not concluding in a general setting.

An other possible way is to consider a partial log-likelihood related to the filtration generated by the observations. If the covariate Z were observable, then the filtration at time t, generated by the observations would be $\sigma\{Z, N(s), \mathbb{1}_{X>s}, 0 \le s \le t \le \tau\}$, and the intensity of the censored process $N(t)$ with respect to this filtration would equal $\lambda(t,\theta^0,Z) =$

$\eta_{\gamma^0}(t)Y(t)f_{\beta^0}(Z)$. In case of covariate measurement error, Z is unobservable and only the evaluation U is available. In this context, the filtration generated by the observations is $\mathcal{E}_t = \sigma\{U, N(s), \mathbb{1}_{X>s}, 0 \le s \le t \le \tau\}$, and the intensity of the censored process $N(t)$ with respect to the filtration $\mathcal{E}_t$ equals

$$\mathbb{E}[\lambda(t,\theta^0,Z)|\mathcal{E}_t] = \eta_{\gamma^0}(t)Y(t)\mathbb{E}[f_{\beta^0}(Z)|\sigma(U,\mathbb{1}_{T\ge t})].$$

This is studied in the Cox model by Prentice (1982) who proposes the following induced failure hazard function

$$\eta_{\gamma^0}(t)\mathbb{E}[\exp(\beta^0 Z)|\sigma(\mathbb{1}_{T\ge t}, U)].$$

The presence of $\{T \ge t\}$ in the conditioning usually implies that the induced partial log-likelihood has not explicit form. When the marginal distribution of Z given $\{T \ge t, U\}$ is specified at each time t, Prentice (1982) proposes an approximation of the induced partial log-likelihood independent of the baseline hazard function. Nevertheless this approximation is appropriate only when the disease is rare. Tsiatis *et al.* (1995) propose another approximation of the induced partial log-likelihood, but emphasise that their method cannot be used for the model of excess relative risk.

In the Cox model with missing covariate Pons (2002) uses also the partial likelihood. She proposes an estimator based on (1.3) where $\mathbb{E}[\exp(\beta Z)|\sigma(U,\mathbb{1}_{T\ge t})]$ is replaced with $\mathbb{E}[\exp(\beta Z)|U]$ by considering that

$$\mathbb{E}[\exp(\beta Z)|\sigma(U,\mathbb{1}_{T\ge t})] = \mathbb{E}[\exp(\beta Z)|U]. \tag{1.7}$$

Nevertheless, obvious examples can be exhibited to prove that the equality (1.7) does not hold in a general setting. As a conclusion, the partial likelihood related to the filtration $\mathcal{E}_t$ seems unusable since it is difficult to separate the estimation of β^0 from the estimation of η_{γ^0}.

1.2. **Our results.** Our estimation procedure is based on the estimation of least squares criterion using deconvolution methods. More precisely, using the observations $\Delta_i = (X_i, D_i, U_i)$ for $i = 1, \dots, n$, we estimate the least squares criterion

$$S_{\theta^0,g}(\theta) = \mathbb{E}\left(f_\beta^2(Z)W(Z)\int_0^\tau Y(t)\eta_\gamma^2(t)dt\right) - 2\mathbb{E}\left(f_\beta(Z)W(Z)\int_0^\tau \eta_\gamma(t)dN(t)\right). \tag{1.8}$$

The function W is a positive weight function to be suitably chosen such that Wf_β and its derivatives up to order 3 with respect to β are in $\mathbb{L}_1(\mathbb{R})\cap\mathbb{L}_2(\mathbb{R})$ and have the best smoothness properties as possible, as functions of z. Under reasonable identifiability assumptions, $S_{\theta^0,g}(\theta)$ is minimum if and only if $\theta = \theta^0$. We propose to estimate $S_{\theta^0,g}(\theta)$ for all $\theta \in \Theta$ by a quantity depending on the observations $\Delta_1, \cdots, \Delta_n$, expecting thus that the argument minimum of the estimator converges to the argument minimum of $S_{\theta^0,g}(\theta)$, say θ^0.

We propose a first estimator of θ^0, say $\widehat{\theta}_1$, constructed by minimizing $S_{n,1}(\theta)$, a consistent estimator of $S_{\theta^0,g}$ where g is replaced by a kernel deconvolution estimator. We show that under classical assumptions, this estimator is a consistent estimator of θ^0. Its rate of convergence depends on the smoothness of f_ε and on the smoothness on $W(z)f_\beta(z)$, as a function of z. More precisely, its depends on the behavior of the ratios of the Fourier transforms $(Wf_\beta)^*(t)/\overline{f_\varepsilon^*}(t)$ and $(Wf_\beta^2)^*(t)/\overline{f_\varepsilon^*}(t)$ as t tends to infinity. We give upper bounds for the risk of $\widehat{\theta}_1$ for various relative risks and various types of error density and derive sufficient conditions ensuring the $\sqrt{n}$-consistency and the asymptotic normality. These upper bounds and these sufficient conditions are deeply studied through examples. In particular we show that $\widehat{\theta}_1$ is a $\sqrt{n}$-consistent asymptotically Gaussian estimator of θ^0 in the Cox model, in the model of excess relative risk, and when f_β is a general polynomial function.

The estimation procedure is related to the problem of the estimation $S_{\theta^0,g}(\theta)$. Under conditions ensuring that it can be estimated at the parametric rate, we propose a second estimator

$\widehat{\theta}_2$ which is $\sqrt{n}$-consistent and asymptotically Gaussian of θ^0. Clearly, these conditions are not always fullfilled and $\widehat{\theta}_2$ does not always exist, whereas $\widehat{\theta}_1$ can be constructed and studied in all setups.

The paper is organized as follows. Section 2 presents the model and the assumptions. In Sections 3 and 4 we present the two estimators and their asymptotic properties illustrated in Section 5. In Section 6, we comment the use of the least squares criterion. The proofs are gathered in Section 7 and in the Appendix.

2. Model, assumptions and notations

Before we describe the estimation procedure, we give notations used throughout the paper and assumptions commonly done in survival data analysis.

Notations For two complex-valued functions u and v in $\mathbb{L}_2(\mathbb{R})\cap\mathbb{L}_1(\mathbb{R})$, let

$$u^*(x)=\int e^{itx}u(t)dt,\quad u\star v(x)=\int u(y)v(x-y)dy,\ \text{and}\quad <u,v>=\int u(x)\overline{v}(x)dx$$

with $\overline{z}$ the conjugate of a complex number z. We also use the notations

$$\|u\|_1=\int|u(x)|dx,\quad \|u\|^2=\int|u(x)|^2dx,\quad \|u\|_\infty=\sup_{x\in\mathbb{R}}|u(x)|,$$

and for $\theta\in\mathbb{R}^d$,

$$\|\ \theta\ \|_{\ell^2}^2=\sum_{k=1}^d\theta_k^2.$$

For a map

$$\begin{array}{rcl}\varphi_\theta & : & \Theta\times\mathbb{R}\longrightarrow\mathbb{R}\\ & & (\theta,u)\mapsto\varphi_\theta(u),\end{array}$$

whenever they exist, the first and second derivatives with respect to θ are denoted by

$$\begin{aligned}\varphi_\theta^{(1)}(\cdot)&=\left(\varphi_{\theta,j}^{(1)}(\cdot)\right)_j\ \text{with}\ \varphi_{\theta,j}^{(1)}(\cdot)=\frac{\partial\varphi_\theta(\cdot)}{\partial\theta_j}\ \text{for}\ j\in\{1,\cdots,m+p\}\\ \text{and}\quad\varphi_\theta^{(2)}(\cdot)&=\left(\varphi_{\theta,j,k}^{(2)}(\cdot)\right)_{j,k}\ \text{with}\ \varphi_{\theta,j,k}^{(2)}(\cdot)=\frac{\partial^2\varphi_\theta(\cdot)}{\partial\theta_j\theta_k},\ \text{for}\ j,k\in\{1,\cdots,m+p\}.\end{aligned}$$

Throughout the paper $\mathbb{P}$, $\mathbb{E}$ and Var denote respectively the probability, the expectation, and the variance when the underlying and unknown true parameters are θ^0 and g. Finally we use the notation a_- for the negative part of a, which equals a if $a\leq 0$ and 0 otherwise.

Model assumptions

($\mathbf{A_1}$) The function η_{γ^0} is non-negative and integrable on $[0,\tau]$.

($\mathbf{A_2}$) Conditionnally on Z and U, the failure time T and the censoring time C are independent.

($\mathbf{A_3}$) The distribution of the censoring time C, conditional on Z and U, does not depend on Z and U.

($\mathbf{A_4}$) The distribution of the failure time T, conditional on Z and U, does not depend on U.

These assumptions are usual in most frameworks dealing with survival data analysis and covariate measured with error, see Andersen *et al.* (1993), Prentice and Self (1983), Prentice(1982), Gong (1990) and Tsiatis (1995). Assumptions ($\mathbf{A_2}$) and ($\mathbf{A_3}$) state that a general censorship model is considered, where the censoring time has an arbitrary distribution independent of the covariates. Assumption ($\mathbf{A_4}$) states that the failure time is independent of the observed covariate when the observed and true covariates are both given, i.e. the measurement error is not prognostic.

We define the filtration

$$\mathcal{F}_t = \sigma\{Z, U, N(s), 1\!\!1_{X\geq s}, 0 \leq s \leq t \leq \tau\}.$$

The intensity of the censored process $N(t)$ with respect to the filtration $\mathcal{F}_t$ equals

$$\lambda(t, \theta^0, Z) = \eta_{\gamma^0}(t)Y(t)f_{\beta^0}(Z). \tag{2.1}$$

It follows from (2.1) and from the independence of the observations Δ_i, that for the individual i the intensity and the compensator process of the censored process $N_i(t) = 1\!\!1_{X_i\leq t, D_i=1}$ with respect to the filtration $\mathcal{F}_t$ are respectively

$$\lambda_i(t, \theta^0, Z_i) = \eta_{\gamma^0}(t)Y_i(t)f_{\beta^0}(Z_i) \text{ and } \Lambda_i(t, \theta^0, Z_i) = \int_0^t \lambda_i(s, \theta^0, Z_i)ds. \tag{2.2}$$

Moreover the process $M_i(t) = N_i(t) - \Lambda_i(t, \theta^0, Z_i)$ is a local square integrable martingale. As a consequence, the least squares criterion defined in (1.8) can be rewritten as

$$\begin{aligned} S_{\theta^0,g}(\theta) = &\int_0^\tau \mathbb{E}\left[\left\{\eta_\gamma(t)f_\beta(Z) - \eta_{\gamma^0}(t)f_{\beta^0}(Z)\right\}^2 Y(t)W(Z)\right] dt \\ &- \int_0^\tau \mathbb{E}\left[\left\{\eta_{\gamma^0}(t)f_{\beta^0}(Z)\right\}^2 Y(t)W(Z)\right] dt. \end{aligned} \tag{2.3}$$

Since we consider general relative risk functions we assume the below minimal smoothness conditions with respect to θ.

Smoothness assumptions

($\mathbf{A_5}$) The functions $\beta \mapsto f_\beta$ and $\gamma \mapsto \eta_\gamma$ admit continuous derivatives up to order 3 with respect to β and γ respectively.

We denote by $S^{(1)}_{\theta^0,g}(\theta)$ and $S^{(2)}_{\theta^0,g}(\theta)$ the first and second derivatives of $S_{\theta^0,g}(\theta)$ with respect to θ. For all t in $[0,\tau]$, set $S^{(2)}_{\theta^0,g}(\theta, t)$ the second derivative of $S_{\theta^0,g}$ when the integral is taken over $[0,t]$, with the convention that $S^{(2)}_{\theta^0,g}(\theta) = S^{(2)}_{\theta^0,g}(\theta, \tau)$.

Identifiability and moment assumptions

($\mathbf{A_6}$) $S^{(1)}_{\theta^0,g}(\theta) = 0$ if and only if $\theta = \theta^0$.

($\mathbf{A_7}$) For all $t \in [0,\tau]$, the matrix $S^{(2)}_{\theta^0,g}(\theta^0, t)$ exists and is positive definite.

($\mathbf{A_8}$) The quantity $\mathbb{E}(f^2_\beta(Z)W(Z))$ is finite .

($\mathbf{A_9}$) For $j = 1, \cdots, m$, $\mathbb{E}|f_{\beta^0}(Z)f^{(1)}_{\beta^0,j}(Z)W(Z)|^3$, $\mathbb{E}|f_{\beta^0}(Z)W(Z)|^3$ are finite.

We denote by $\mathcal{G}$ the set of densities g such that the assumptions ($\mathbf{A_2}$)-($\mathbf{A_4}$),($\mathbf{A_6}$)-($\mathbf{A_9}$) hold.

3. Construction and study of the first estimator $\widehat{\theta}_1$

3.1. **Construction.** If the Z_i's were observed, $S_{\theta^0,g}(\theta)$ would be estimated by

$$\tilde{S}_n(\theta) = -\frac{2}{n}\sum_{i=1}^n f_\beta(Z_i)W(Z_i)\int_0^\tau \eta_\gamma(t)dN_i(t) + \frac{1}{n}\sum_{i=1}^n f^2_\beta(Z_i)W(Z_i)\int_0^\tau \eta^2_\gamma(t)Y_i(t)dt \tag{3.1}$$

and θ^0 would be estimated by minimizing $\tilde{S}_n(\theta)$. Since the Z_i's are unobservable and Z_i independent of ε_i, the density h of U_i equals $h = g \star f_\varepsilon$. We thus estimate $S_{\theta^0,g}$ by

(3.2)
$$S_{n,1}(\theta) = -\frac{2}{n}\sum_{i=1}^{n}(f_\beta W)\star K_{n,C_n}(U_i)\int_0^\tau \eta_\gamma(t)dN_i(t) + \frac{1}{n}\sum_{i=1}^{n}(f_\beta^2 W)\star K_{n,C_n}(U_i)\int_0^\tau \eta_\gamma^2(t)Y_i(t)dt,$$

where $K_{n,C_n}(\cdot) = C_n K_n(C_n\cdot)$ is a deconvolution kernel defined via its Fourier transform, such that $\int K_n(x)dx = 1$, and

$$K^*_{n,C_n}(t) = \frac{K^*_{C_n}(t)}{\overline{f^*_\varepsilon}(t)} = \frac{K^*(t/C_n)}{\overline{f^*_\varepsilon}(t)}, \tag{3.3}$$

with K^* compactly supported satisfying $|1 - K^*(t)| \le \mathbb{1}_{|t|\ge 1}$ and $C_n \to \infty$ as $n \to \infty$.

Using this criterion we propose to estimate θ^0 by

$$\widehat{\theta}_1 = \begin{pmatrix}\widehat{\beta}_1\\ \widehat{\gamma}_1\end{pmatrix} = \arg\min_{\theta=(\beta,\gamma)^\top\in\Theta} S_{n,1}(\theta). \tag{3.4}$$

We require for the construction of $S_{n,1}(\theta)$, that

($\mathbf{A_{10}}$) the density f_ε belongs to $\mathbb{L}_2(\mathbb{R})\cap\mathbb{L}_\infty(\mathbb{R})$ and for all $x\in\mathbb{R}$, $f^*_\varepsilon(x)\neq 0$.

3.2. **Asymptotic properties of the first $\widehat{\theta}_1$.** Assume that

($\mathbf{A_{11}}$) $\sup_{g\in\mathcal{G}} \| f^2_{\beta^0} g \|_2^2 \le C_2(f^2_{\beta^0})$, $\sup_{g\in\mathcal{G}} \| f_{\beta^0} g \|_2^2 \le C_2(f_{\beta^0})$.

($\mathbf{A_{12}}$) $\sup_{\beta\in\mathbb{B}}(Wf_\beta)$, W and $\sup_{\beta\in\mathbb{B}}(Wf^2_\beta)$ belong to $\mathbb{L}_1(\mathbb{R})$.

($\mathbf{A_{13}}$) $\sup_{\beta\in\mathbb{B}}(Wf^{(1)}_\beta)$ and $\sup_{\beta\in\mathbb{B}}(Wf_\beta f^{(1)}_\beta)$ belong to $\mathbb{L}_1(\mathbb{R})$.

As in density deconvolution, or for the estimation of the regression function in errors-in-variables models, the rate of convergence for estimating θ^0 is given by both the smoothness of f_ε and the smoothness of $(f_\beta W)(z)$, and $\partial(f_\beta W)(z)/\partial\beta$, as functions of z. The smoothness of the error density f_ε is described by the decrease of its Fourier transform.

($\mathbf{A_{14}}$) There exist positive constants $\underline{C}(f_\varepsilon), \overline{C}(f_\varepsilon)$, and nonnegative δ, ρ, α and u_0 such that $\underline{C}(f_\varepsilon) \le |f^*_\varepsilon(u)|\,|u|^\alpha \exp(\delta\,|u|^\rho) \le \overline{C}(f_\varepsilon)$ for all $|u|\ge u_0$.

If $\rho = 0$, by convention $\delta = 0$. When $\rho = 0 = \delta$ in ($\mathbf{A_{14}}$), f_ε is called "ordinary smooth". When $\delta > 0$ and $\rho > 0$, it is called "super smooth". Densities satisfying ($\mathbf{A_{14}}$) with $\rho > 0$ and $\delta > 0$ are infinitely differentiable. The standard examples for super smooth densities are the Gaussian or Cauchy distributions which are super smooth of respective order $\alpha = 0, \rho = 2$ and $\alpha = 0, \rho = 1$. For ordinary smooth densities, one can cite for instance the double exponential (also called Laplace) distribution with $\rho = 0 = \delta$ and $\alpha = 2$. We consider here that $0 \le \rho \le 2$. The square integrability of f_ε in ($\mathbf{A_{10}}$) requires that $\alpha > 1/2$ when $\rho = 0$ in ($\mathbf{A_{14}}$).

The smothness of $f_\beta W$ is described by the following assumption.

($\mathbf{A_{15}}$) There exist positive constants $\underline{L}(f), \overline{L}(f)$ and a, d, u_0, r nonnegative numbers such that for all $\beta\in\mathbb{B}, f_\beta W$ and $f^2_\beta W$ and their derivatives up to order 3 with respect to β, belong to

$$\mathcal{H}_{a,d,r} = \{f\in\mathbb{L}_1(\mathbb{R}); \underline{L}(f) \le |f^*(u)||u|^a\exp(d|u|^r) \le \overline{L}(f) < \infty \text{ for all } |u|\ge u_0\}. \tag{3.5}$$

If $r = 0$, by convention $d = 0$.

Theorem 3.1. *Let* $(\mathbf{A_1})$*-*$(\mathbf{A_{15}})$ *hold. Let* $\widehat{\theta}_1 = \widehat{\theta}_1(C_n)$ *be defined by (3.2) and (3.4) with* C_n *a sequence such that*

$$C_n^{(2\alpha-2a+1-\rho+(1-\rho)_-)}\exp\{-2dC_n^r+2\delta C_n^\rho\}/n = o(1) \text{ as } n\to+\infty. \tag{3.6}$$

1) Then $\mathbb{E}(\| \widehat{\theta}_1(C_n)-\theta^0\|^2_{\ell^2}) = o(1)$*, as* $n\to\infty$ *and* $\widehat{\theta}_1(C_n)$ *is a consistent estimator of* θ^0*.*
2) Moreover, $\mathbb{E}(\| \widehat{\theta}_1-\theta^0\|^2_{\ell^2}) = O(\varphi_n^2)$ *with* $\varphi_n^2 = \|(\varphi_{n,j})\|^2_{\ell^2}$*,* $\varphi_{n,j}^2 = B_{n,j}^2(\theta^0)+V_{n,j}(\theta^0)/n$*, where* $B^2_{n,j}(\theta^0) = \min\{B^{[1]}_{n,j}(\theta^0), B^{[2]}_{n,j}(\theta^0)\}$*,* $V_{n,j}(\theta^0) = \min\{V^{[1]}_{n,j}(\theta^0), V^{[2]}_{n,j}(\theta^0)\}$*, with*

$$\begin{aligned}
B^{[q]}_{n,j}(\theta^0) &= \left\|(f^2_{\beta^0}W)^*(K^*_{C_n}-1)\right\|^2_q + \left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|^2_q + \left\|\left(f^{(1)}_{\beta^0,j}W\right)^*(K^*_{C_n}-1)\right\|^2_q\\
&\quad + \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|^2_q,\\
V^{[q]}_{n,j}(\theta^0) &= \left\|(f^2_{\beta^0}W)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|^2_q + \left\|(f_{\beta^0}W)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|^2_q + \left\|\left(f^{(1)}_{\beta^0,j}W\right)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|^2_q\\
&\quad + \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|^2_q.
\end{aligned}$$

3) Furthermore, the resulting rate φ_n^2 *is given in Table 1*

We point out that the rate for estimating β^0 depends on the smoothness properties of $\partial(Wf_\beta)(z)/\partial\beta$ and $\partial(Wf^2_\beta)(z)/\partial\beta$ as a function of z, whereas, the rate for estimating γ^0 depends on the smoothness properties of $Wf_\beta(z)$ and $Wf^2_\beta(z)$ as a function of z. In both cases, the smoothness properties of η_γ as a function of t does not have influence on the rate of convergence.

The terms $B^2_{n,j}$ and $V_{n,j}$ are respectively the squared bias and variance terms. As usual ,the bias is the smallest for the smoothest functions $(Wf_\beta)(z)$ and $\partial(f_\beta W)(z)/\partial\beta$, as functions of z. As in density deconvolution, the biggest variance are obtained for the smoothest error density f_ε. Hence, the slowest rates are obtained for the smoothest errors density f_ε, for instance for Gaussian ε's.

The rate of convergence of the estimator $\widehat{\theta}_1$ could be improved by assuming smoothness properties on the density g. But, since g is unknown, we choose to not assume such properties. Consequently, without any additional assumptions on g, the parametric rate of convergence is achieved as soon as (Wf_β) and (Wf^2_β) and their derivatives, as functions of z, are smoother than the errors density f_ε.

3.3. **Consequence : a sufficient condition to obtain the parametric rate of convergence with $\widehat{\theta}_1$.**

$(\mathbf{C_1})$ There exists a weight function W such that the functions $\sup_{\beta\in\mathbb{B}}(f_\beta W)^*/\overline{f^*_\varepsilon}$, $\sup_{\beta\in\mathbb{B}}(f^2_\beta W)^*/\overline{f^*_\varepsilon}$ belong to $\mathbb{L}_1(\mathbb{R})\cap\mathbb{L}_2(\mathbb{R})$.

$(\mathbf{C_2})$ The functions $\sup_{\beta\in\mathbb{B}}\left(f^{(1)}_\beta W\right)^*/\overline{f^*_\varepsilon}$ and $\sup_{\beta\in\mathbb{B}}\left(f^{(1)}_\beta f_\beta W\right)^*/\overline{f^*_\varepsilon}$ belong to $\mathbb{L}_1(\mathbb{R})\cap\mathbb{L}_2(\mathbb{R})$.

$(\mathbf{C_3})$ The functions $\left(f^{(2)}_\beta W\right)^*/\overline{f^*_\varepsilon}$ and $\left(\dfrac{\partial^2(f^2_\beta W)}{\partial\beta^2}\right)^*/\overline{f^*_\varepsilon}$ belong to $\mathbb{L}_1(\mathbb{R})\cap\mathbb{L}_2(\mathbb{R})$ for all $\beta\in\mathbb{B}$.

		f_ε: $\rho = 0$ in $(\mathbf{A_{14}})$ ordinary smooth	f_ε: $\rho > 0$ in $(\mathbf{A_{14}})$ super smooth
Wf_{β^0}	$d = r = 0$ in $(\mathbf{A_{15}})$ Sobolev	$a < \alpha + 1/2$ $\quad n^{-\frac{2a-1}{2\alpha}}$	$(\log n)^{-\frac{2a-1}{\rho}}$
		$a \geq \alpha + 1/2$ $\quad n^{-1}$	
	$r > 0$ in $(\mathbf{A_{15}})$ $\mathcal{C}^\infty$	n^{-1}	$r < \rho$ $\quad (\log n)^{A(a,r,\rho)} \exp\left\{-2d\left(\frac{\log n}{2\delta}\right)^{r/\rho}\right\}$
			$r = \rho$, $d < \delta$ $\quad (\log n)^{A(a,r,\rho)+2\alpha d/(\delta r)} n^{-d/\delta}$
			$r = \rho$, $d = \delta,\ a < \alpha + 1/2$ $\quad (\log n)^{(2\alpha-2a+1)/r} n^{-1}$
			$r = \rho$, $d = \delta,\ a \geq \alpha + 1/2$ $\quad n^{-1}$
			$r = \rho$, $d > \delta$ $\quad n^{-1}$
			$r > \rho$ $\quad n^{-1}$

where $A(a, r, \rho) = (-2a + 1 - r + (1 - r)_-)/\rho$.

TABLE 1. Rates of convergence φ_n^2 of $\widehat{\theta}_1$

Theorem 3.2. *Let* $(\mathbf{A_1})$*-*$(\mathbf{A_{12}})$ *and* $(\mathbf{C_1})$*-*$(\mathbf{C_3})$ *hold. Then* $\widehat{\theta}_1$ *defined by (3.4) is a* $\sqrt{n}$*-consistent estimator of* θ^0*. Moreover*

$$\sqrt{n}(\widehat{\theta}_1 - \theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_1),$$

where Σ_1 *equals*

(3.7)
$$\left(\mathbb{E}\left[-2\int_0^\tau \frac{\partial^2((f_\beta W)(Z)\eta_\gamma(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\, dN(s) + \int_0^\tau \frac{\partial^2((f_\beta^2 W)(Z)\eta_\gamma^2(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\, Y(s)ds\right]\right)^{-1}$$
$$\times\Sigma_{0,1}\left(\mathbb{E}\left[-2\int_0^\tau \frac{\partial^2((f_\beta W)(Z)\eta_\gamma(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\, dN(s) + \int_0^\tau \frac{\partial^2((f_\beta^2 W)(Z)\eta_\gamma^2(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\, Y(s)ds\right]\right)^{-1}$$

with

$$\Sigma_{0,1} = \mathbb{E}\left\{\left[-2\int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta}\,|_{\theta=\theta^0}\, dN(s) + \int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta}\,|_{\theta=\theta^0}\, Y(s)ds\right]\right.$$
$$\left.\times\left[-2\int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta}\,|_{\theta=\theta^0}\, dN(s) + \int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta}\,|_{\theta=\theta^0}\, Y(s)ds\right]^\top\right\}$$

where

$$R_{\beta,f_\varepsilon,1}(U)=\int (Wf_\beta)^*(t)\frac{e^{-itU}}{\overline{f_\varepsilon^*(t)}}dt \quad \textit{and} \quad R_{\beta,f_\varepsilon,2}(U)=\int (Wf_\beta^2)^*(t)\frac{e^{-itU}}{\overline{f_\varepsilon^*(t)}}dt.$$

The conditions **(C$_1$)-(C$_3$)**, stronger than **(A$_{14}$)** and **(A$_{15}$)**, ensure the existence of the functions $R_{\beta,f_\varepsilon,j}$ for $j=1,2$.

4. Construction and study of the second estimator $\widehat{\theta}_2$

4.1. **Construction.** Our estimation procedure, based on the estimation of the least squares criterion, requires the estimation of $\mathbb{E}[\int_0^\tau (f_\beta W)(Z)dN(t)]$ and $\mathbb{E}[\int_0^\tau (f_\beta^2 W)(Z)Y(t)dt]$, which are linear functional of g. It may appear that these linear functional could be directly estimated, without kernel deconvolution plugg-in. In this context, we propose another estimator of θ^0. It is based on sufficient conditions allowing to construct a $\sqrt{n}$-consistent estimator of these linear functionals and hence to estimate $S_{\theta^0,g}$ with the parametric rate.

We say that the conditions (**C$_4$**)-(**C$_6$**) hold if there exist a weight function W and two functions $\Phi_{\beta,f_\varepsilon,1}$ and $\Phi_{\beta,f_\varepsilon,2}$ not depending on g, such that for all $\beta\in\mathbb{B}$ and for all g

$$(\mathbf{C_4}) \qquad \mathbb{E}_{\theta^0,g}\left[\int_0^\tau (f_\beta W)(Z)dN(t)\right]=\mathbb{E}_{\theta^0,h}\left[\int_0^\tau \Phi_{\beta,f_\varepsilon,1}(U)dN(t)\right]$$
$$\text{and } \mathbb{E}_{\theta^0,g}\left[\int_0^\tau (f_\beta^2 W)(Z)Y(t)dt\right]=\mathbb{E}_{\theta^0,h}\left[\int_0^\tau \Phi_{\beta,f_\varepsilon,2}(U)Y(t)dt\right];$$

$$(\mathbf{C_5}) \qquad \text{For } k=0,1,2 \text{ and for } j=1,2, \quad \mathbb{E}[\sup_{\beta\in\mathbb{B}}\|\Phi^{(k)}_{\beta,f_\varepsilon,j}(U)\|_{\ell^2}]<\infty;$$

$$(\mathbf{C_6}) \qquad \text{For } j=1,2 \text{ and for all } \beta\in\mathbb{B},\ \mathbb{E}\left[\|\Phi^{(1)}_{\beta,f_\varepsilon,j}(U)\|^2_{\ell^2}\right]<\infty.$$

Under (**C$_4$**)-(**C$_6$**), we estimate $S_{\theta^0 g}$ by

$$S_{n,2}(\theta)=-\frac{2}{n}\sum_{i=1}^n\int_0^\tau \Phi_{\beta,f\varepsilon,1}(U_i)\eta_\gamma(t)dN_i(t)+\frac{1}{n}\sum_{i=1}^n\int_0^\tau \Phi_{\beta,f\varepsilon,2}(U_i)\eta_\gamma^2(t)Y_i(t)dt \tag{4.1}$$

and θ^0 is estimated by

$$\widehat{\theta}_2=\arg\min_{\theta\in\Theta}S_{n,2}(\theta). \tag{4.2}$$

The main difficulty for finding such functions $\Phi_{\beta,f_\varepsilon,1}$ and $\Phi_{\beta,f_\varepsilon,2}$ lies in the constraint that they must not depend on the unknown density g. We refer to Section 4.3 for details on how to construct such functions $\Phi_{\beta,f_\varepsilon,j}$, $j=1,2$.

4.2. Asymptotic properties of $\widehat{\theta}_2$.

Theorem 4.1. *Let (**A$_1$**)-(**A$_9$**), and the conditions (**C$_4$**)-(**C$_6$**) hold. Then $\widehat{\theta}_2$, defined by (4.2) is a $\sqrt{n}$-consistent estimator of θ^0. Moreover*

$$\sqrt{n}(\widehat{\theta}_2-\theta^0)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\Sigma_2),$$

where Σ_2 *equals*

(4.3)

$$\left(\mathbb{E}\left[-2\int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,dN(s)+\int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,Y(s)ds\right]\right)^{-1}$$
$$\times\Sigma_{0,2}\left(\mathbb{E}\left[-2\int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,dN(s)+\int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,Y(s)ds\right]\right)^{-1}$$

with

$$\Sigma_{0,2}=\mathbb{E}\left\{\left[-2\int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta}\,|_{\theta=\theta^0}\,dN(s)+\int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta}\,|_{\theta=\theta^0}\,Y(s)ds\right]\right.$$
$$\left.\times\left[-2\int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta}\,|_{\theta=\theta^0}\,dN(s)+\int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta}\,|_{\theta=\theta^0}\,Y(s)ds\right]^\top\right\}.$$

4.3. Comments on conditions ensuring $\sqrt{n}$-consistency : comparison of $\widehat{\theta}_1$ and $\widehat{\theta}_2$.

Let us briefly compare the conditions ($\mathbf{C_1}$)-($\mathbf{C_3}$) to the conditions ($\mathbf{C_4}$)-($\mathbf{C_6}$). It is noteworthy that the conditions ($\mathbf{C_4}$)-($\mathbf{C_6}$) are more general. First, the condition ($\mathbf{C_4}$) does not require that $f_\beta W$, $f_\beta^2 W$ belong to $\mathbb{L}_1(\mathbb{R})$ (as for instance in the Cox Model). Second, we point out that Condition ($\mathbf{C_1}$) implies ($\mathbf{C_4}$), with $\Phi_{\beta,f_\varepsilon,j}=R_{\beta,f_\varepsilon,j}$. This comes from the facts that under ($\mathbf{C_1}$)-($\mathbf{C_3}$), by denoting $\Phi_{\beta,f_\varepsilon,1}=(Wf_\beta)^*/\overline{f_\varepsilon^*}$ and $\Phi_{\beta,f_\varepsilon,2}=(Wf_\beta^2)^*/\overline{f_\varepsilon^*}$, we have

$$\begin{aligned}
\mathbb{E}[Y(t)\Phi_{\beta,f_\varepsilon,2}(U)] &= \iiint \mathbb{1}_{x\geq t}\Phi_{\beta,f_\varepsilon,2}(u)f_{X,Z}(x,z)f_\varepsilon(u-z)dx\,du\,dz\\
&= \iint \mathbb{1}_{x\geq t}f_{X,Z}(x,z)\frac{1}{2\pi}\int \Phi^*_{\beta,f_\varepsilon,2}(s)e^{-isz}\overline{f_\varepsilon^*}(s)ds\,dx\,dz\\
&= \iint \mathbb{1}_{x\geq t}f_{X,Z}(x,z)\frac{1}{2\pi}\int \frac{(Wf_\beta^2)^*(s)}{\overline{f_\varepsilon^*}(s)}e^{-isz}\overline{f_\varepsilon^*}(s)ds\,dx\,dz\\
&= \mathbb{E}[Y(t)(Wf_\beta^2)(Z)].
\end{aligned}$$

Consequently

$$\mathbb{E}\Big[\int_0^\tau \Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(t)Y(t)dt\Big] = \mathbb{E}\Big[\int_0^\tau f_\beta^2(Z)W(Z)\eta_\gamma^2(t)dt\Big],$$

and analogoulsy

$$\mathbb{E}\Big[\int_0^\tau \Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(t)dN(t)\Big] = \mathbb{E}\Big[\int_0^\tau f_\beta(Z)W(Z)\eta_\gamma(t)dN(t)\Big].$$

Hence Condition ($\mathbf{C_4}$) holds and $\Sigma_{0,1}=\Sigma_{0,2}$ with $\Sigma_{0,1}$ defined in Theorem 3.2.

These comments underline the key importance of the weight function W. For instance, if $f_\beta(z)=1-\beta+\beta/(1+z^2)$, and f_ε is the Gaussian density, then it seems impossible to find a function $\Phi_{\beta,f_\varepsilon,2}$ such that $\mathbb{E}[Y(t)\Phi_{\beta,f_\varepsilon,2}(U)]=\mathbb{E}[Y(t)f_\beta^2(Z)]$, whereas ($\mathbf{C_1}$)-($\mathbf{C_3}$) hold by taking $W(z)=(1+z^2)^4\exp(-z^2/(4\delta))$. In this special example, we exhibit a suitable choice of W that ensures that condition ($\mathbf{C_1}$)-($\mathbf{C_3}$) are fulfilled (see Section 5 for further details). Nevertheless, such weight function are not always available and hence those conditions ($\mathbf{C_1}$)-($\mathbf{C_3}$) are not always fulfilled.

5. Examples

In this section, we illustrate the asymptotic properties of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ for various relative risks and error density f_ε. In all of these examples, $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and the noise distribution is arbitrary, as far as it satisfies $(\mathbf{A_{10}})$ and $(\mathbf{A_{14}})$ with $0 \leq \rho \leq 2$.

The first example deals with Cox model. We show that our estimation procedure, based on a nonparametric method and specifically on density deconvolution, also provides $\sqrt{n}$-consistent and asymptotically Gaussian estimator of β^0. The aim of this example is to show that we recover previous known results using estimators that are quite different from the ones proposed by Nakamura (1992) and studied by Kong and Gu (1999) or from the ones proposed by Augustin (2004).

The others examples we consider, deal with relative risks for which no consistent estimators were known when the covariate is mismeasured.

Example 1. Exponential relative risk (Cox model)

Let f_β be of the form $f_\beta(z) = \exp(\beta z)$ and assume that $\mathbb{E}[\exp(2\beta^0 U)] < \infty$. Let $W(z) = \exp\{-z^2/(4\delta)\}$. Then the conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as the condition $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

One could also choose $W \equiv 1$ and use that $\mathbb{E}[\exp(\beta Z)] = \mathbb{E}[\exp(\beta U)]/\mathbb{E}[\exp(\beta\varepsilon)]$. This implies that if we denote by

$$\Phi_{\beta,f_\varepsilon,1}(U) = \frac{\exp(2\beta U)}{\mathbb{E}[\exp(2\beta\varepsilon)]} \text{ and } \Phi_{\beta,f_\varepsilon,2}(U) = \frac{\exp(\beta U)}{\mathbb{E}[\exp(\beta\varepsilon)]}$$

then $\mathbb{E}[\Phi_{\beta,f_\varepsilon,1}(U)] = \mathbb{E}[f_\beta^2(Z)]$ and $\mathbb{E}[Y(t)\Phi_{\beta,f_\varepsilon,2}(U)] = \mathbb{E}[Y(t)f_\beta(Z)]$, and the criterion $S_{n,2}$ defined by (4.1) exists.

In this case $\widehat{\theta}_2$ is also a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

Example 2. Polynomial relative risk 1 (included Excess relative risk model) Let f_β be of the form $f_\beta(z) = 1 + \sum_{k=1}^m \beta_k z^k$ and let $W(z) = \exp\{-z^2/(4\delta)\}$. Then conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

We point out that when $m = 1$, $f_\beta(z) = 1 + \beta z$, and this model is known as the model of excess relative risk.

One can also choose $W \equiv 1$, provided that the kernel K has finite absolute moments of order m and satisfies $\int u^r K(u)du = 0$, for $r = 1, \ldots, m$. With this choice of W, $\widehat{\theta}_1$ remains a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

Example 3. Cosines relative risk 1 Let f_β be of the form $f_\beta(z) = \sum_{j=1}^m \beta_j \cos(jz)$ with $\sum_{j=1}^m \beta_j = 1$. Let $W(z) = \exp\{-z^2/(4\delta)\}$. Then the conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

One can also choose $W \equiv 1$ and use that $\mathbb{E}[\exp(ijZ)] = \mathbb{E}[\exp(ijU)]/\mathbb{E}[\exp(ij\varepsilon)]$. This implies that if we denote by

$$\Phi_{\beta,f_\varepsilon,1}(U) = \frac{1}{2}\left[\frac{\exp(ijU)}{f_\varepsilon^*(j)} + \frac{\exp(-ijU)}{\overline{f_\varepsilon^*}(j)}\right]$$

and

$$\Phi_{\beta,f_\varepsilon,2}(U) = \frac{1}{4}\Bigg\{1 + \sum_{j=1}^{m}\beta_j^2\left[\frac{\exp(2ijU)}{f_\varepsilon^*(2j)} + \frac{\exp(-2ijU)}{\overline{f_\varepsilon^*}(2j)}\right]$$

$$+\sum_{j=1}^{m}\sum_{k\neq j}\beta_j\beta_k\left[\frac{\exp(i(j+k)U)}{f_\varepsilon^*(j+k)} + \frac{\exp(-i(j+k)U)}{\overline{f_\varepsilon^*}(j+k)} + \frac{\exp(i(j-k)U)}{f_\varepsilon^*(j-k)} + \frac{\exp(i(-j+k)U)}{\overline{f_\varepsilon^*}(j-k)}\right]\Bigg\}$$

then the criterion $S_{n,2}$ defined in (4.1) exists.

With this choice of W, $\widehat{\theta}_2$ remains a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0. In the same way, $\widehat{\theta}_1$ with $W \equiv 1$ also remains $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

Example 4. Cauchy relative risk 1 Consider f_β of the form $f_\beta(z) = 1-\beta+\beta/(1+z^2)$. Then f_β has the regularity of $z \mapsto 1/(1+z^2)$ which belongs to $\mathcal{H}_{a,d,r}$ defined in **($\mathbf{A_{15}}$)** with $a=0$, $d=1/2$ and $r=1$. Let $W(z) = (1+z^2)^4\exp\{-z^2/(4\delta)\}$. Hence the functions $f_\beta W$, $f_\beta^2 W$ and their derivatives in β up to order 3 belong to $\mathcal{H}_{a,d,r}$ defined in **($\mathbf{A_{15}}$)** with $\rho < r = 2$ or $\rho = r = 2$ and $d > \delta$. Consequently, the conditions ($\mathbf{C_1}$)-($\mathbf{C_3}$) as well as conditions ($\mathbf{C_4}$)-($\mathbf{C_6}$) are satisfied and the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

This simple example underlines the importance of the smoothing weight function W in the construction of $\widehat{\theta}_1$ or $\widehat{\theta}_2$. Indeed, without a smoothing function W in front of the relative risk, Theorem 3.1 predicts a rate of convergence of order $\exp(-2\sqrt{\log n})$ for Gaussian ε.

Example 5. Laplace relative risk Consider f_β of the form $f_\beta(z) = 1+\beta f(z)$ with $f(z) = \exp(-|z|/2)-1$. Since the Fourier transform of $z \mapsto \exp(-|z|/2)$ is slowly decaying, like $|u|^{-2}$ as $|u| \to \infty$, if we choose $W \equiv 1$, the estimator $\widehat{\theta}_1$ is not $\sqrt{n}$-consistent as soon as $|f_\varepsilon^*(u)| \le o(|u|^{-2})$ with $|u| \to \infty$. A closer look tells us that f_β and its derivative in β is $\mathcal{C}^\infty$ except at one point $z=0$. Therefore, a proper choice of W can smooth out at 0 and make Wf_β, Wf_β^2 and their derivatives in β infinitely differentiable functions in z. This choice of W ensures the $\sqrt{n}$-consistency of $\widehat{\theta}_1$ whatever f_ε satisfies **($\mathbf{A_{14}}$)** with $0<\rho<1$. Even if $\rho \ge 1$, the rate of $\widehat{\theta}_1$ is much faster when using our choice of W then it would be for $W \equiv 1$. Let us precise the choice of W. Set

$$\Psi_{A,B,R}(z) = \exp\left(-\frac{1}{(z-A)^R(B-z)^R}\right) I_{[A,B]}(z), \tag{5.1}$$

where $-\infty < A < B < \infty$ are fixed and $R>0$. According to Lepski and Levit (1998) and Fedoryuk (1987), p. 346, Theorem 7.3, $|\Psi_{A,B,R}^*(u)| \le c\exp(-C|u|^{R/(R+1)})$, as $|u| \to \infty$ and c, C are positive constants. We propose to take W equal to $\Psi_{0,100,R}$ or $\Psi_{-100,0,R}$ or their sum.

This choice of W ensures that $f_\beta W$, $f_\beta^2 W$ and their derivatives up to order 3 belong to $\mathcal{H}_{a,d,r}$ defined in **($\mathbf{A_{15}}$)** with $d>0$ and $r = R/(R+1)$ closer to 1 as R comes larger.

If f_ε satisfies **($\mathbf{A_{14}}$)** with $0 \le \rho < 1$, we choose R large enough such that $r = R/(R+1) > \rho$. Hence, the conditions ($\mathbf{C_1}$)-($\mathbf{C_3}$) as well as the conditions ($\mathbf{C_4}$)-($\mathbf{C_6}$) are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho \geq 1$, for this choice of W, the functions Wf_β and Wf_β^2 and their derivatives in β up to order 3, belong to $\mathcal{H}_{a,d,r}$ with $r = R/(R+1)$ and hence, according to Table 1,

$$\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1)\,(\log n)^{\frac{1-2a-r}{\rho}} \exp\{-2d(\log n/(2\delta))^{r/\rho}\}.$$

Example 6. Irregular relative risk Consider f_β of the form $f_\beta(z) = 1 - \beta + \beta \mathbb{1}_{[-1,1]}(z)$ and take $W = \Psi_{-1,1,R}$ defined by (5.1) for $R > 0$.

If $\rho = 0$ in $(\mathbf{A_{14}})$, then Conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as Conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho > 0$, then the best rate for estimating θ^0 is obtained by choosing $W = \Psi_{-1,1,R}$ with $R > 0$ sufficiently large such that Wf_β and Wf_β^2 and their derivatives in β up to order 3, belong to $\mathcal{H}_{a,d,r}$ defined in $(\mathbf{A_{15}})$ with $0 < r = R/(R+1) < 1$ as close to 1 as needed.

It follows that if $0 \leq \rho < 1$, then we can find $W = \Psi_{-1,1,R}$ belonging to $\mathcal{H}_{a,d,r}$ with $r = R/(R+1) > \rho$. Hence the conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied and the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0.

If $\rho \geq 1$, for $W = \Psi_{-1,1,R}$, the functions Wf_β and Wf_β^2 and their derivatives in β up to order 3, belong to $\mathcal{H}_{a,d,r}$ with $r = R/(R+1)$ and hence, according to Table 1,

$$\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1)\,(\log n)^{\frac{1-2a-r}{\rho}} \exp\{-2d(\log n/(2\delta))^{r/\rho}\}.$$

Example 7. Polygonal relative risk Consider f_β with $f_\beta(z) = 1 - \beta_2 a_- - \beta_3|b|^3 + \beta_1 z + \beta_2(z-a)\mathbb{1}_{z\geq a} + \beta_3|z-b|^3$. This relative risk is $\mathcal{C}^\infty$ except at points a and b where it is not differentiable. We suggest to use the smoothing weight function in (5.1) as follows. For $R > 0$, let

$$W(z) = \Psi_{a-100,a,R}(z) + \Psi_{a,b,R}(z) + \Psi_{b,b+100,R}(z).$$

If the noise satisfies $(\mathbf{A_{14}})$ with $0 \leq \rho < 1$, then take R large enough such that $r = R/(R+1) > \rho$ and thus conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho \geq 1$ in $(\mathbf{A_{14}})$, the functions Wf_β and Wf_β^2 and their derivatives in β up to order 3, belong to $\mathcal{H}_{a,d,r}$ with $r = R/(R+1)$ and hence, according to Table 1

$$\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1)\,(\log n)^{\frac{1-2a-r}{\rho}} \exp\{-2d(\log n/(2\delta))^{r/\rho}\}.$$

Comments on the examples 5, 6, and 7

In these three examples, $f_\beta W$ belongs to $\mathcal{H}_{a,d,r}$ defined in $(\mathbf{A_{15}})$ with r at most such that $r < 1$. Hence $\widehat{\theta}_1$ achieves the $\sqrt{n}$-rate of convergence provided that f_ε is ordinary smooth or super smooth with an exponent $\rho < 1$. It seems therefore impossible to have $(Wf_\beta)^*/f_\varepsilon^*$ in $\mathbb{L}_1(\mathbb{R})$ when the ε_i's are Gaussian. This comes from the fact that for these relative risks, the least squares criterion $S_{\theta^0,g}(\theta)$ cannot be estimated with the parametric rate of convergence and hence could probably, not provide a $\sqrt{n}$-consistent estimator of θ^0. Nevertheless, even in cases where the parametric rate of convergence seems not achievable by such estimators, the resulting rate of the risk of $\widehat{\theta}_1$ is clearly infinitely faster than the logarithmic rate predicted by Table 1 that we could have without W .

In most of previous examples where the weight function W is required, the points where $f_\beta(z)$ has to be smoothed do not depend on β. But in survival data analysis the relative risks

f_β are usually of the form $f_\beta(z) = f(\beta z)$ (see for instance Prentice and Self (1983)). In such models, the points where $f_\beta(z)$ has to be smoothed (as function of z) will depend on β.

Let us present such examples.

Example 8. Polynomial relative risk 2 Let f_β be of the form $f(\beta z)$ with $f(z) = 1 + \sum_{k=1}^m a_k z^k$ with known a_k's. Let $W(z) = \exp\{-z^2/(4\delta)\}$. Then conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Example 9. Cosines relative risk 2 Let f_β be of the form $f(\beta z)$ with $f(z) = \sum_{j=1}^m a_j \cos(jz)$ with known a_k's such that $\sum_{j=1}^m a_j = 1$. Let $W(z) = \exp\{-z^2/(4\delta)\}$. Then the conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Example 10. Cauchy relative risk 2 Consider f_β of the form $f(\beta z)$ with $f(z) = 1/(1+z^2)$. Let $W(z) = (1 + z^2)^4 \exp\{-z^2/(4\delta)\}$ or $W \equiv 1$. With these choices of W, the functions $f_\beta W$, $f_\beta^2 W$ and their derivatives in β up to order 3 belong to $\mathcal{H}_{a,d,r}$ defined in $(\mathbf{A_{15}})$ with $a = 0$, $d = 1/\beta$ and $r = 1$. According to Table 1, if f_ε satisfies $(\mathbf{A_{14}})$ with $0 \leq \rho < 1$, then $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance. If f_ε satisfies $(\mathbf{A_{14}})$ with $\rho \geq 1$ and then $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are consistent with a rate that depends on β^0. Let us be more precise. According to the proof of Theorem 3.1, for $j = 1, \ldots, m+p$, the term $B_{n,j}^2(\theta^0)$ are of order $\exp(-2C_n/\beta^0)$ and the term $V_{n,j}(\theta^0)/n$ are of order $C_n^{2\alpha+(1-\rho)+(1-\rho_-)} \exp(-2C_n/\beta^0 + 2\delta C_n^\rho)/n$.

Set C_n^* that realizes the best compromise between the squared bias and the variance terms. It is independent from β^0 and is given by

$$C_n^* = \left[\frac{\log n}{2\delta} - \frac{(2\alpha + (1-\rho)_-)}{2\delta\rho} \log\left(\frac{\log n}{2\delta}\right)\right]^{1/\rho}.$$

This choice yields to the rate

$$\varphi_n^2 = \max\left\{n^{-1}, \exp\left[-\frac{2}{\beta^0}\left(\frac{\log n}{2\delta} - \frac{2\alpha + (1-\rho)_-}{2\delta\rho} \log\left(\frac{\log n}{2\delta}\right)\right)^{1/\rho}\right] (\log n)^{(1-\rho)/\rho}\right\}.$$

In other words, if $\rho = 1$, then $\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1) \max\left\{n^{-1}, n^{-1/(\beta^0\delta)}(\log n)^{2\alpha/(\beta^0\delta)}\right\}$ and if $\rho > 1$, then $\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1) \exp\left[-2(\beta^0)^{-1} (\log n/(2\delta))^{1/\rho}\right]$.

6. Comment on the use of the least squares criterion

In a proportional hazard model without errors, the main drawback of the least squares criterion, compared to the partial log-likelihood, is that it does not allow to separate the estimation of β^0 from the estimation of the baseline hazard function η. The subject of this part is to motivate the choice of the least squares criterion when the covariate is mismeasured.

First, as it is mentionned in the introduction, the partial log-likelihood related to the filtration given by the observations only has not an explicit form.

Second, consider as a partial log-likelihood related to the failure hazard function defined by (1.1)

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \log\left\{\frac{\mathbb{E}(f_\beta(Z_i)|\sigma(U_i, \mathbb{1}_{T_i \geq t})}{n^{-1}\sum_{j=1}^n Y_j(t)\mathbb{E}[f_\beta(Z_j)|\sigma(U_j, \mathbb{1}_{T_j \geq t})]}\right\} dN_i(t).$$

This partial log-likelihood depends on the observations, on the density g of Z and on η_{γ^0}, through $\mathbb{E}[f_\beta(Z)|\sigma(U, \mathbb{1}_{T\geq t})]$. Hence the estimation of β^0 also depends on η through the conditionning.

Lastly, since the Z_i's are unobservable, one other idea would be to estimate β^0 by minimizing $\hat{L}_n^{(1)}(\beta, U^{(n)})$ given by

$$\hat{L}_n^{(1)}(\beta, U^{(n)}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left[\left(\frac{f_\beta^{(1)}W}{f_\beta}\right)\star K_{n,C_n}(U_i) - \frac{\sum_{j=1}^n Y_j(t)(f_\beta^{(1)}W)\star K_{n,C_n}(U_j)}{\sum_{j=1}^n Y_j(t)(f_\beta W)\star K_{n,C_n}(U_j)}\right] W\star K_{n,C_n}(U_i)dN_i(t),$$

for W a suitable chosen weight function, with $W(z)\neq 0$ for all z in $\mathbb{R}$. Due to the unobservability of $Z^{(n)}$, $\hat{L}_n^{(1)}(\beta, U^{(n)})$ can be seen as an estimation of the expectation of (1.3). Under reasonnable assumptions, $\hat{\beta}^{PL}$ such that $\hat{L}_n^{(1)}(\hat{\beta}^{PL}, U^{(n)}) = 0$ is a consistent estimator of β^0. The main difficulty lies in the study of its rate of convergence. As in the study of $\widehat{\theta}_1$, the rate of convergence of $\hat{\beta}^{PL}$ depends on the smoothness of $(f_\beta^{(1)}W)(z)/(f_\beta)(z)$, as a function of z, through the behavior of the ratio

$$\frac{\left((f_{\beta^0}^{(1)}W)/f_{\beta^0}\right)^*(t)}{\overline{f_\varepsilon^*}(t)}, \text{ as } t\to\infty.$$

Consequently, the best properties would be obtained for W such that $(f_\beta^{(1)}W)/(f_\beta)$ is in $\mathbb{L}_1(\mathbb{R})$ and has the best smoothness properties. In the Cox model, $f_\beta^{(1)}(z)/f_\beta(z) = z$ and this estimation criterion provides $\sqrt{n}$-consistency and asymptotically Gaussian estimator, analogously to the Nakamura's (1992)'s estimator. The same result holds for the relative risks considered in Examples 3 and 9. Nevertheless, for general relative risks, this criterion is less tractable than the least squares criterion (3.2), since it is strictly more difficult to "smooth" $z\mapsto f_\beta^{(1)}(z)/f_\beta(z)$ than $z\mapsto f_\beta(z)$. This appears in a crucial way in the model of excess relative risk where $f_\beta^{(1)}(z)/f_\beta(z) = z/(1+\beta z)$. This point has to be related to the difficulty and even the impossibility to find a suitable correction of $L_n^{(1)}(\beta, U^{(n)})$, which leads to asymptotically unbiased score functions (see the introduction).

7. Proofs

From now C denotes any numerical constant and $C(A)$ indicates that it depends on a A.

7.1. **Proof of Theorem 3.1.**

7.1.1. ***Consistency***. By classical arguments, the consistency follows from the two points :

1- The quantity $S_{\theta^0,g}(\theta)$ is minimum if and only if $\theta = \theta^0$.

2- For all $\theta\in\Theta$, $\mathbb{E}[S_{n,1}(\theta) - S_{\theta^0,g}(\theta)]^2 = o(1)$ as $n\to\infty$, with $S_{\theta^0,g}(\theta)$ defined in (1.8),

3- If $\omega(n,\rho)$ denotes $\omega(n,\rho) = \sup\{|S_{n,1}(\theta) - S_{n,1}(\theta')| : \|\theta-\theta'\|_{\ell^2}\leq\rho\}$, there exists ρ_k tending to 0, such that $\mathbb{E}[\omega(n,\rho_k)]^2 = O(\rho_k^2)$ as $n\to\infty$ $\forall k\in\mathbb{N}$.

Proof of 1- Under $(\mathbf{A_6})$, by applying (2.3) we get

$$\frac{\partial}{\partial\beta}S_{\theta^0,g}(\theta) = 2\int_0^\tau \mathbb{E}\left[f_\beta^{(1)}(Z_i)\eta_\gamma(t)\left\{\eta_\gamma(t)f_\beta(Z) - \eta_{\gamma^0}(t)f_{\beta^0}(Z)\right\}W(Z)Y(t)\right]dt = 0 \Leftrightarrow \theta = \theta^0,$$

and

$$\frac{\partial}{\partial\gamma}S_{\theta^0,g}(\theta) = 2\int_0^\tau \mathbb{E}\left[f_\beta(Z_i)\eta_\gamma^{(1)}(t)\left\{\eta_\gamma(t)f_\beta(Z) - \eta_{\gamma^0}(t)f_{\beta^0}(Z)\right\}W(Z)Y(t)\right]dt = 0 \Leftrightarrow \theta = \theta^0.$$

The matrix of second derivatives equals

$$\left(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta^2}\,|_{\theta=\theta^0}\right) \quad := \quad H(\theta^0) = \begin{pmatrix} H_{11}(\theta^0) & H_{12}(\theta^0) \\ (H_{12}(\theta^0))^\top & H_{22}(\theta^0)\end{pmatrix}$$

with

$$\begin{aligned} H_{11}(\theta^0) &= 2\int_0^\tau \mathbb{E}\left[(f_{\beta^0}^{(1)}(Z))(f_{\beta^0}^{(1)}(Z))^\top \eta_{\gamma^0}^2(t)W(Z)Y(t)\right]dt \\ H_{12}(\theta^0) &= 2\int_0^\tau \mathbb{E}\left[f_{\beta^0}(Z)\eta_{\gamma^0}(t)f_{\beta^0}^{(1)}(Z)(\eta_{\gamma^0}^{(1)}(t))^\top W(Z)Y(t)\right]dt \\ H_{22}(\theta^0) &= 2\int_0^\tau \mathbb{E}\left[(\eta_{\gamma^0}^{(1)}(t))(\eta_{\gamma^0}^{(1)}(t))^\top f_{\beta^0}^2(Z)W(Z)Y(t)\right]dt. \end{aligned}$$

An obvious application of Cauchy-Schwarz Inequality gives that the matrix H is non negative definite and hence under $(\mathbf{A_7})$ **1-** is proved.

Proof of 2-

For both the bias and the variance, we will give two upper bounds, based on the two following applications of the Hölder's inequality

$$|<\varphi_1,\varphi_2>| \leq \|\varphi_1\|_2\|\varphi_2\|_2, \tag{7.1}$$

and

$$|<\varphi_1,\varphi_2>| \leq \|\varphi_1\|_\infty\|\varphi_2\|_1\,. \tag{7.2}$$

According to Lemma 8.1 we write that

$$\mathbb{E}[S_{n,1}(\theta)] = \int_0^\tau \mathbb{E}\left[(f_\beta^2 W)\star K_{C_n}(Z)\eta_\gamma^2(t)Y(t) - 2(f_\beta W)\star K_{C_n}(Z)\eta_\gamma(t)f_{\beta^0}(Z)\eta_{\gamma^0}(t)Y(t)\right]dt,$$

and hence

$$\begin{aligned} \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta) &= \int\int_0^\tau \eta_\gamma^2(t)\mathbb{1}_{x\geq t}\left\langle f_{X,Z}(x,\cdot),(f_\beta^2W)\star K_{C_n} - f_\beta^2 W\right\rangle dx\,dt \\ &\quad -2\int\int_0^\tau \eta_{\gamma^0}(t)\eta_\gamma(t)\mathbb{1}_{x\geq t}\left\langle f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot),(f_\beta W)\star K_{C_n} - f_\beta W\right\rangle dx\,dt. \end{aligned}$$

By applying (7.1) we obtain the first bound

$$\begin{aligned} \left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right| &\leq \Big(\int_0^\tau \eta_\gamma^2(t)dt\Big)\,\|f_{X,Z}\|_2\|(f_\beta^2W)\star K_{C_n} - (f_\beta^2 W)\|_2 \\ &\quad + \Big(2\int_0^\tau \eta_\gamma(t)\eta_{\gamma^0}(t)dt\Big)\int \|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_2\,dx\,\|(f_\beta W)\star K_{C_n} - f_\beta W\|_2\,. \end{aligned}$$

Applying Parseval's formula we get

$$\begin{aligned} \left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right| &\leq (2\pi)^{-1}\Big(\int_0^\tau \eta_\gamma^2(t)dt\Big)\,\|f_{X,Z}\|_2\|(f_\beta^2W)^*(K_{C_n}^* - 1)\|_2 \\ &\quad + (\pi)^{-1}\Big(\int_0^\tau \eta_\gamma(t)\eta_{\gamma^0}(t)dt\Big)\int \|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_2\,dx\,\|(f_\beta W)^*(K_{C_n}^* - 1)\|_2 \end{aligned}$$

that is

$$\begin{aligned} \left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right| &\leq C(\gamma,\gamma^0,f_{\beta^0}) \\ &\quad \left[\|(f_\beta^2W)^*(K_{C_n}^* - 1)\|_2 + \|(f_\beta W)^*(K_{C_n}^* - 1)\|_2\right]. \end{aligned} \tag{7.3}$$

According to (7.2), $\left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right|$ is also bounded by

$$\begin{aligned}
&\| (f_\beta^2 W)\star K_{C_n} - (f_\beta^2 W) \|_\infty \int \| f_{X,Z}(x,\cdot) \|_1 \, dx \Big(\int_0^\tau \eta_\gamma^2(t)dt \Big) \\
&\quad + \| (f_\beta W)\star K_{C_n} - (f_\beta W) \|_\infty \Big(2 \int \| f_{\beta^0}(\cdot) f_{X,Z}(x,\cdot) \|_1 \, dx \int_0^\tau \eta_\gamma(t)\eta_{\gamma^0}(t)dt \Big) \\
&\qquad \le \| (f_\beta^2 W)^*(K^*_{C_n} - 1) \|_1 \Big((2\pi)^{-1} \| f_{X,Z} \|_1 \int_0^\tau \eta_\gamma^2(t)dt \Big) \\
&\quad + \| (f_\beta W)^*(K^*_{C_n} - 1) \|_1 \Big(\pi^{-1} \int \| f_{\beta^0}(\cdot) f_{X,Z}(x,\cdot) \|_1 \, dx \int_0^\tau \eta_\gamma(t) f_{\beta^0}\eta_{\gamma^0}(t)dt \Big).
\end{aligned}$$

This implies that

$$\begin{aligned}
\left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right| \;\le\; & \Big[\int_0^\tau \eta_\gamma^2(t)dt \Big] \| (f_\beta^2 W)^*(K^*_{C_n} - 1) \|_1 \\
& + \Big[\mathbb{E}|f_{\beta^0}(Z)| \int_0^\tau \eta_\gamma(t)\eta_{\gamma^0}(t)dt \Big] \| (f_\beta W)^*(K^*_{C_n} - 1) \|_1,
\end{aligned}$$

that is

$$\begin{aligned}
\left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right| \le\; & C(\gamma,\gamma^0,f_{\beta^0}) \\
& \left[\| (f_\beta W)^*(K^*_{C_n} - 1) \|_1 + \| (f_\beta^2 W)^*(K^*_{C_n} - 1) \|_1 \right]. \qquad (7.4)
\end{aligned}$$

By combining the bounds (7.3) and (7.4) we get that

$$\begin{aligned}
\left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right| \le\; & C(\gamma,\gamma^0,f_{\beta^0}) \\
& \times \min \left\{ \| (f_\beta W)^*(K^*_{C_n} - 1) \|_2 + \| (f_\beta^2 W)^*(K^*_{C_n} - 1) \|_2, \right. \\
& \left. \| (f_\beta W)^*(K^*_{C_n} - 1) \|_1 + \| (f_\beta^2 W)^*(K^*_{C_n} - 1) \|_1 \right\}. \qquad (7.5)
\end{aligned}$$

By applying Lemma 8.2

$$\left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right|^2 = O\Big(C_n^{-2a+1-r+(1-r)_-} \exp(-2dC_n^r) \Big) = o(1).$$

Study of the variance Since the random variables are i.i.d., we get that

$$\mathrm{Var}[S_{n,1}(\theta)] \;=\; \frac{(2+o(1))}{n}(A_1 + A_2),$$

with

$$\begin{aligned}
A_1 &= \mathbb{E}\Big[(f_\beta^2 W)\star K_{n,C_n}(U) \int_0^\tau \eta_\gamma^2(t) Y(t)dt \Big]^2 \\
\text{and } A_2 &= 4\mathbb{E}\Big[(f_\beta W)\star K_{n,C_n}(U) \int_0^\tau \eta_\gamma(t) dN(t) \Big]^2.
\end{aligned}$$

According to (7.2) and by applying Lemma 8.1, A_1 is less than

$$\begin{aligned}
&\Big(\int_0^\tau \eta_\gamma^2(t)dt \Big)^2 \int \left| \left\langle f_{X,Z}(x,\cdot)\star f_\varepsilon, ((f_\beta^2 W)\star K_{n,C_n})^2 \right\rangle \right| dx \\
&\qquad \le \Big(\int_0^\tau \eta_\gamma^2(t)dt \Big)^2 \int \| f_{X,Z}(x,\cdot)\star f_\varepsilon \|_\infty \, dx \, \| (f_\beta^2 W)\star K_{n,C_n} \|_2^2
\end{aligned}$$

and hence

$$A_1 \;\le\; (2\pi)^{-1} \Big(\int_0^\tau \eta_\gamma^2(t)dt \Big)^2 \| f_\varepsilon \|_\infty \| f_{X,Z} \|_1 \left\| \frac{(f_\beta^2 W)^* K^*_{C_n}}{\overline{f^*_\varepsilon}} \right\|_2^2 .$$

In the same way, we get a first bound for A_2. Let us denote by

$$\varphi(X,Z) = \int_0^\tau \eta_\gamma(t)dN(t). \tag{7.6}$$

According to Lemma 8.1 and to (7.2), A_2 is bounded by

$$4\int \left\langle (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon, ((f_\beta W)\star K_{n,C_n})^2\right\rangle dx$$
$$\leq 4\int \| (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \|_\infty \, dx \, \| (f_\beta W)\star K_{n,C_n} \|_2^2 .$$

Since

$$\int \| \varphi^2(x,\cdot)f_{X,Z}(x,\cdot) \|_1 \, dx = \mathbb{E}\Big[\int_0^\tau \eta_\gamma(t)dN(t)\Big]^2,$$

we get that

$$\int \| \varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \|_\infty \, dx \leq \| f_\varepsilon \|_\infty \, \mathbb{E}\Big[\int_0^\tau \eta_\gamma(t)dN(t)\Big]^2.$$

Consequently,

$$A_2 \leq 4\left[(2\pi)^{-1}\mathbb{E}\Big(\int_0^\tau \eta_\gamma(t)dN(t)\Big)^2 \| f_\varepsilon \|_\infty\right] \left\|\frac{(f_\beta W)^* K_{C_n}^*}{\overline{f_\varepsilon^*}}\right\|_2^2.$$

It follows that,

$$\mathrm{Var}[S_{n,1}(\theta)] \leq \frac{C(\theta^0, \| f_\varepsilon \|_\infty)}{n}\Big[\Big\|(f_\beta W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_2^2 + \Big\|(f_\beta^2 W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_2^2\Big]. \tag{7.7}$$

According to (7.2), A_1 is also less than

$$\Big(\int_0^\tau \eta_\gamma^2(t)dt\Big)^2 \int \left|\left\langle f_{X,Z}(x,\cdot)\star f_\varepsilon, ((f_\beta^2 W)\star K_{n,C_n})^2\right\rangle\right| dx$$
$$\leq \Big(\int_0^\tau \eta_\gamma^2(t)dt\Big)^2 \, \| \int f_{X,Z}(x,\cdot) \|_1 \, dx \, \| (f_\beta^2 W)\star K_{n,C_n} \|_\infty^2 .$$

In the same way A_2 is less than

$$4\int \left\langle (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon, ((f_\beta W)\star K_{n,C_n})^2\right\rangle dx$$
$$\leq 4\int \| (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \|_1 \, dx \, \| (f_\beta W)\star K_{n,C_n} \|_\infty^2$$

where $\varphi(X,Z)$ is defined in (7.6). Once again, since

$$\int \| (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \|_1 \, dx = \mathbb{E}\Big(\int_0^\tau \eta_\gamma(t)dN(t)\Big)^2,$$

$$\mathrm{Var}[S_{n,1}(\theta)] \leq \frac{C(\theta^0)}{n}\Big[\Big\|(f_\beta W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_1^2 + \Big\|(f_\beta^2 W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_1^2\Big]. \tag{7.8}$$

By combining (7.7) and (7.8), we obtain that

$$\mathrm{Var}[S_{n,1}(\theta)] \leq \frac{C(\theta^0, \| f_\varepsilon \|_\infty)}{n} \min\Big\{\Big\|(f_\beta W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_2^2 + \Big\|(f_\beta^2 W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_2^2, \Big\|(f_\beta W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_1^2 + \Big\|(f_\beta^2 W)^*\frac{K_{C_n}^*}{\overline{f_\varepsilon^*}}\Big\|_1^2\Big\}. \tag{7.9}$$

According to Lemma 8.2

$$\mathrm{Var}[S_{n,1}(\theta)] = O\Big(C_n^{2(\alpha-a)+1-\rho+(1-\rho)_-}\exp(-2dC_n^r + 2\delta C_n^\rho)/n\Big),$$

and hence under (3.6), $\mathbb{E}\left[S_{n,1}(\theta) - S_{\theta^0,g}(\theta)\right]^2 = o(1)$, as $n\to\infty$.

Proof of 3-

By definition $S_{n,1}(\theta) - S_{n,1}(\theta')$ equals

$$-\frac{2}{n}\int_0^\tau \left[(f_\beta W)\star K_{n,C_n}(U_i)\eta_\gamma(t) - (f_{\beta'}W)\star K_{n,C_n}(U_i)\eta_{\gamma'}(t)\right] dN_i(t)$$
$$+\frac{1}{n}\int_0^\tau \left[(f_\beta^2 W)\star K_{n,C_n}(U_i)\eta_\gamma^2(t) - (f_{\beta'}^2 W)\star K_{n,C_n}(U_i)\eta_{\gamma'}^2(t)\right] Y_i(t)dt.$$

Under $(\mathbf{A_5})$, $(\mathbf{A_{12}})$, $(\mathbf{A_{13}})$, $(\mathbf{A_{14}})$ and $(\mathbf{A_{15}})$, for C_n satisfying (3.6), since $\|\theta-\theta'\|_{\ell^2}\le\rho_k$, we get that $\mathbb{E}(|S_{n,1}(\theta) - S_{n,1}(\theta')|^2) = O(\rho_k^2)$. Hence **3-** follows. □

7.1.2. **Rate of convergence.** Denote by $S_{n,1}^{(1)}(\theta)$ and $S_{n,1}^{(2)}(\theta)$ the first and second derivatives of $S_{n,1}(\theta)$ with respect to θ. By using classical Taylor expansion and the consistency of $\widehat{\theta}_1$, we get that $0 = S_{n,1}^{(1)}(\widehat{\theta}_1) = S_{n,1}^{(1)}(\theta^0) + S_{n,1}^{(2)}(\theta^0)(\widehat{\theta}_1-\theta^0) + R_n(\widehat{\theta}_1-\theta^0)$, with R_n defined by

$$R_n = \int_0^1 [S_{n,1}^{(2)}(\theta^0 + s(\widehat{\theta}_1-\theta^0)) - S_{n,1}^{(2)}(\theta^0)]ds. \tag{7.10}$$

This implies that

$$\widehat{\theta}_1 - \theta^0 = -[S_{n,1}^{(2)}(\theta^0) + R_n]^{-1}S_{n,1}^{(1)}(\theta^0). \tag{7.11}$$

Consequently we have to check the three following points

i) $\mathbb{E}\Big[\big\{S_{n,1}^{(1)}(\theta^0)) - S_{\theta^0,g}^{(1)}(\theta^0)\big\}\big\{S_{n,1}^{(1)}(\theta^0)) - S_{\theta^0,g}^{(1)}(\theta^0)\big\}^\top\Big] = O[\varphi_n\varphi_n^\top]$,

ii) $\mathbb{E}\left[S_{n,1}^{(2)}(\theta^0) - S_{\theta^0,g}^{(2)}(\theta^0)\right]^2 = o(1)$,

iii) R_n defined in (7.10) satisfies $\mathbb{E}(\| R_n\|_{\ell^2}^2) = o(1)$ as $n\to\infty$.

iv) $\mathbb{E}\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^2 \le 4\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top\Big[\Big(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\theta_k}|_{\theta=\theta^0}\Big)^{-1}\Big]^\top\Big(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\theta_k}|_{\theta=\theta^0}\Big)^{-1}S_{n,1}^{(1)}(\theta^0)\right] + o(\varphi_n^2)$.

The rate of convergence of $\widehat{\theta}_1$ is thus given by the order of $S_{n,1}^{(1)}(\theta^0) - S_{\theta^0,g}^{(1)}(\theta^0) = S_{n,1}^{(1)}(\theta^0)$.

Proof of i)

According to (3.2), $S_{n,1}^{(1)}(\theta^0)$ equals

(7.12)
$$\frac{2}{n}\sum_{i=1}^n\left(\begin{array}{c} -\int_0^\tau (f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(t)dN_i(t) + \int_0^\tau (f_{\beta^0}f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}^2(t)Y_i(t)dt \\ -\int_0^\tau (f_{\beta^0}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}^{(1)}(t)dN_i(t) + \int_0^\tau (f_{\beta^0}^2W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(t)\eta_{\gamma^0}^{(1)}(t)Y_i(t)dt\end{array}\right).$$

Study of the bias By definition, $\mathbb{E}(\partial S_{n,1}(\theta)/\partial\beta)_{\theta=\theta^0}$ equals

$$-2\mathbb{E}\Big[\int_0^\tau (f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_1)\eta_{\gamma^0}(t)dN_1(t)\Big] + 2\mathbb{E}\Big[\int_0^\tau (f_{\beta^0}f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_1)\eta_{\gamma^0}^2(t)Y_1(t)dt\Big].$$

Hence, according to Lemma 8.1,

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\beta}\,|_{\theta=\theta^0}\right) &= -2\mathbb{E}\left[f_{\beta^0}(Z_1)(f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right] \\
&\quad +2\mathbb{E}\left[(f_{\beta^0}f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right] \\
&= -2\mathbb{E}\left[f_{\beta^0}(Z_1)(f^{(1)}_{\beta^0}W)\star K_{C_n}(Z_1)\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right] \\
&\quad +2\mathbb{E}\left[(f_{\beta^0}f^{(1)}_{\beta^0}W)\star K_{C_n}(Z_1)\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right].
\end{aligned}$$

Since

$$\left.\frac{\partial S_{\theta^0,g}(\theta)}{\partial\beta}\right|_{\theta=\theta^0} = -2\mathbb{E}\left[f_{\beta^0}(Z_1)(f^{(1)}_{\beta^0}W)(Z_1)\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right]+2\mathbb{E}\left[(f_{\beta^0}f^{(1)}_{\beta^0}W)(Z_1)\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right] = 0,$$

we get that $\mathbb{E}\left(\partial S_{n,1}(\theta)/\partial\beta|_{\theta=\theta^0}\right)$ also equals

$$\begin{aligned}
&-2\mathbb{E}\left[f_{\beta^0}(Z_1)[(f^{(1)}_{\beta^0}W)\star K_{C_n}(Z_1)-(f^{(1)}_{\beta^0}W)(Z_1)]\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right] \\
&\qquad +2\mathbb{E}\left[[(f_{\beta^0}f^{(1)}_{\beta^0}W)\star K_{C_n}(Z_1)-(f_{\beta^0}f^{(1)}_{\beta^0}W)(Z_1)]\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right] \\
&\quad = -2\int\left\langle f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot),[(f^{(1)}_{\beta^0}W)\star K_{C_n}-(f^{(1)}_{\beta^0}W)]\right\rangle\int_0^\tau \eta^2_{\gamma^0}(t)\mathbb{1}_{x\geq t}dt\,dx \\
&\qquad\quad +2\int\left\langle f_{X,Z}(x,\cdot),[(f_{\beta^0}f^{(1)}_{\beta^0}W)\star K_{C_n}-(f_{\beta^0}f^{(1)}_{\beta^0}W)]\right\rangle\int_0^\tau \eta^2_{\gamma^0}(t)\mathbb{1}_{x\geq t}dt\,dx.
\end{aligned}$$

In the same way, according to Lemma 8.1,

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\gamma}\,|_{\theta=\theta^0}\right) &= 2\mathbb{E}\left[-(f_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta^{(1)}_{\gamma^0}(t)dN_1(t)\right] \\
&\quad +2\mathbb{E}\left[(f^2_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta_{\gamma^0}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right] \\
&= -2\mathbb{E}\left[f_{\beta^0}(Z_1)(f_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right] \\
&\quad +2\mathbb{E}\left[(f^2_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right] \\
&= -2\mathbb{E}\left[f_{\beta^0}(Z_1)(f_{\beta^0}W)\star K_{C_n}(Z_1)\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right] \\
&\quad +2\mathbb{E}\left[(f^2_{\beta^0}W)\star K_{C_n}(Z_1)\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right].
\end{aligned}$$

Since

$$\begin{aligned}
\frac{\partial S_{\theta^0,g}(\theta)}{\partial\gamma_j}\,|_{\theta=\theta^0} &= -2\mathbb{E}\left[f_{\beta^0}(Z_1)(f_{\beta^0}W)(Z_1)\int_0^\tau \eta^{(1)}_{\gamma^0}\eta_{\gamma^0}(t)Y_1(t)dt\right] \\
&\quad +2\mathbb{E}\left[(f^2_{\beta^0}W)(Z_1)\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right] \\
&= 0,
\end{aligned}$$

we obtain that $\mathbb{E}\left(\partial S_{n,1}(\theta)/\partial\gamma\ |_{\theta=\theta^0}\right)$ equals

$$\begin{aligned}
&- \quad 2\mathbb{E}\left[f_{\beta^0}(Z_1)[(f_{\beta^0}W)\star K_{C_n}(Z_1)-(f_{\beta^0}W)(Z_1)]\int_0^\tau \eta^{(1)}_{\gamma^0}\eta_{\gamma^0}(t)Y_1(t)dt\right]\\
&\quad +2\mathbb{E}\left[[(f^2_{\beta^0}W)\star K_{C_n}(Z_1)-(f^2_{\beta^0}W)(Z_1)]\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right]\\
&= -2\int\left\langle f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot),[(f_{\beta^0}W)\star K_{C_n}-(f_{\beta^0}W)]\right\rangle\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)\mathbb{1}_{x\geq t}dt\,dx\\
&\quad +2\int\left\langle f_{X,Z}(x,\cdot),[(f^2_{\beta^0}W)\star K_{C_n}-(f^2_{\beta^0}W)]\right\rangle\int_0^\tau \eta^{(1)}_{\gamma^0}(t)\eta_{\gamma^0}(t)\mathbb{1}_{x\geq t}dt\,dx.
\end{aligned}$$

A first bound for this bias term can be obtained by writing that for $j=1,\cdots,m$

$$(1/2)\mathbb{E}\left(\partial S_{n,1}(\theta)/\partial\beta_j|_{\theta=\theta^0}\right)$$

is bounded by

$$\begin{aligned}
&\parallel\left(f^{(1)}_{\beta^0,j}W\right)\star K_{C_n}-\left(f^{(1)}_{\beta^0,j}W\right)\parallel_2\left(\int\parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau \eta^2_{\gamma^0}(t)dt\right)\\
&\quad+\left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)\star K_{C_n}-\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)\right\|_2\left(\int\parallel f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau \eta^2_{\gamma^0}(t)dt\right)\\
&\quad\leq\left\|\left(f^{(1)}_{\beta^0,j}W\right)^*(K^*_{C_n}-1)\right\|_2\left((2\pi)^{-1}\int\parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau \eta^2_{\gamma^0}(t)dt\right)\\
&\qquad+\left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|_2\left((2\pi)^{-1}\int\parallel f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau \eta^2_{\gamma^0}(t)dt\right).
\end{aligned}$$

In the same way,

$$(1/2)\mathbb{E}\left(\partial S_{n,1}(\theta)/\partial\gamma_j|_{\theta=\theta^0}\right)$$

is bounded by

$$\begin{aligned}
&\parallel(f_{\beta^0}W)\star K_{C_n}-f_{\beta^0}W\parallel_2\left(\int\parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau\left|\eta^{(1)}_{\gamma^0,j}(t)\right|\eta_{\gamma^0}(t)dt\right)\\
&\quad+\parallel(f^2_{\beta^0}W)\star K_{C_n}-f^2_{\beta^0}W\parallel_2\left(\int\parallel f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau\left|\eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)\right|dt\right)\\
&\leq\parallel(f_{\beta^0}W)^*(K^*_{C_n}-1)\parallel_2\left((2\pi)^{-1}\int\parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau\left|\eta^{(1)}_{\gamma^0,j}(t)\right|\eta_{\gamma^0}(t)dt\right)\\
&\quad+\parallel(f^2_{\beta^0}W)^*(K^*_{C_n}-1)\parallel_2\left((2\pi)^{-1}\int\parallel f_{X,Z}(x,\cdot)\parallel_2 dx\int_0^\tau\left|\eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)\right|dt\right).
\end{aligned}$$

Consequently

$$\begin{aligned}
(7.13)\quad &\left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\beta_j}\ |_{\theta=\theta^0}\right)\right|\\
&\qquad\leq C(\theta^0)\left[\left\|\left(f^{(1)}_{\beta^0,j}W\right)^*(K^*_{C_n}-1)\right\|_2+\left\|\left((f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|_2\right],
\end{aligned}$$

and

$$\begin{aligned}
(7.14)\quad &\left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\gamma_j}\ |_{\theta=\theta^0}\right)-\left(\frac{\partial S_{\theta^0,g}(\theta)}{\partial\gamma_j}\ |_{\theta=\theta^0}\right)\right|\\
&\qquad\leq C(\theta^0)\left[\left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_2+\left\|\left(f^2_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|_2\right].
\end{aligned}$$

A second bound for the bias term can be obtained by writing that $(1/2)\mathbb{E}\left(\partial S_{n,1}(\theta)/\partial\beta_j \mid_{\theta=\theta^0}\right)$ is bounded by

$$\begin{aligned}
&\parallel \left(f^{(1)}_{\beta^0,j}W\right)\star K_{C_n} - \left(f^{(1)}_{\beta^0,j}W\right) \parallel_\infty \left(\int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^2_{\gamma^0}(t)dt\right)\\
&\quad + \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)\star K_{C_n} - \left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)\right\|_\infty \left(\int \parallel f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^2_{\gamma^0}(t)dt\right)\\
&\quad \le \left\|\left(f^{(1)}_{\beta^0,j}W\right)^*(K^*_{C_n}-1)\right\|_1 \left((2\pi)^{-1}\int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^2_{\gamma^0}(t)dt\right)\\
&\qquad + \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|_1 \left((2\pi)^{-1}\int \parallel f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^2_{\gamma^0}(t)dt\right).
\end{aligned}$$

In the same way $(1/2)\left|\mathbb{E}\left(\partial S_{n,1}(\theta)/\partial\gamma_j \mid_{\theta=\theta^0}\right)\right|$ is bounded by

$$\begin{aligned}
&\parallel (f_{\beta^0}W)\star K_{C_n} - f_{\beta^0}W \parallel_\infty \left(\int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)dt\right)\\
&\quad + \parallel (f^2_{\beta^0}W)\star K_{C_n} - f^2_{\beta^0}W \parallel_\infty \left(\int \parallel f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)dt\right)\\
&\le \parallel (f_{\beta^0}W)^*(K^*_{C_n}-1)\parallel_1 \left((2\pi)^{-1}\int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)dt\right)\\
&\quad + \parallel (f^2_{\beta^0}W)^*(K^*_{C_n}-1)\parallel_1 \left((2\pi)^{-1}\int \parallel f_{X,Z}(x,\cdot)\parallel_1 dx \int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)dt\right).
\end{aligned}$$

Consequently

$$\begin{aligned}
(7.15)\quad &\left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\beta_j}\mid_{\theta=\theta^0}\right)\right|\\
&\qquad \le C(\theta^0)\left[\left\|\left(f^{(1)}_{\beta^0,j}W\right)^*(K^*_{C_n}-1)\right\|_1 + \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|_1\right],
\end{aligned}$$

and

$$(7.16)\quad \left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\gamma_j}\mid_{\theta=\theta^0}\right)\right| \le C(\theta^0)\left[\left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1 + \left\|(f^2_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1\right].$$

By combining (7.13), (7.14), (7.15) and (7.16) we get that

$$\begin{aligned}
(7.17)\quad &\left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\beta_j}\mid_{\theta=\theta^0}\right)\right| \le C(\theta^0)\\
&\qquad \times \min\left\{\left\|\left(f^{(1)}_{\beta^0,j}W\right)^*(K^*_{C_n}-1)\right\|_2^2 + \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|_2^2,\right.\\
&\qquad\qquad \left.\left\|\left(f^{(1)}_{\beta^0,j}W\right)^*(K^*_{C_n}-1)\right\|_1^2 + \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*(K^*_{C_n}-1)\right\|_1^2\right\}
\end{aligned}$$

and

$$\begin{aligned}
(7.18)\quad &\left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial\gamma_j}\mid_{\theta=\theta^0}\right)\right| \le C(\theta^0)\\
&\qquad \times \min\left\{\left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_2^2 + \left\|(f^2_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_2^2,\right.\\
&\qquad\qquad \left.\left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1^2 + \left\|(f^2_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1^2\right\}.
\end{aligned}$$

According to Lemma 8.2

$$\left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial \beta_j}\mid_{\theta=\theta^0}\right)\right|^2 = O\left(C_n^{-2a+1-r+(1-r)_-}\exp(-2dC_n^r)\right),$$

and

$$\left|\mathbb{E}\left(\frac{\partial S_{n,1}(\theta)}{\partial \gamma_j}\mid_{\theta=\theta^0}\right)\right|^2 = O\left(C_n^{-2a+1-r+(1-r)_-}\exp(-2dC_n^r)\right).$$

Study of the variance

For the variance term, it is easy to see that

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial \beta_j}\mid_{\theta=\theta^0}\right) = \frac{8+o(1)}{n}\left[V_{1,j}+V_{2,j}\right],$$

with

$$V_{1,j} = \mathbb{E}\left[\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)\star K_{n,C_n}(U_1)\int_0^\tau \eta^2_{\gamma^0}(t)Y_1(t)dt\right]^2$$

and

$$V_{2,j} = \mathbb{E}\left[\left(f^{(1)}_{\beta^0,j}W\right)\star K_{n,C_n}(U_1)\int_0^\tau \eta_{\gamma^0}(t)dN_1(t)\right]^2$$

In the same way

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial \gamma_j}\right) = \frac{8+o(1)}{n}\left[V_{3,j}+V_{4,j}\right],$$

with

$$V_{3,j} = \mathbb{E}\left[(f^2_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)Y_1(t)dt\right]^2$$

and

$$V_{4,j} = \mathbb{E}\left[(f_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)dN_1(t)\right]^2.$$

According to Lemma 8.1,

$$V_{1,j} \le \left[\int_0^\tau \eta^2_{\gamma^0}(t)dt\right]^2\int\left|\left\langle f_{X,Z}(x,\cdot)\star f_\varepsilon, \left(\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)\star K_{n,C_n}\right)^2\right\rangle\right|dx$$

and

$$V_{3,j} \le \left[\int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)dt\right]^2\int\left|\left\langle f_{X,Z}(x,\cdot)\star f_\varepsilon, \left((f^2_{\beta^0}W)\star K_{n,C_n}\right)^2\right\rangle\right|dx.$$

By applying the inequalities (7.1) and (7.2) we get that

$$\begin{aligned} V_{1,j} \le\ & \left[\int_0^\tau \eta^2_{\gamma^0}(t)dt\right]^2 \\ & \times\min\left\{\parallel f_\varepsilon\parallel_\infty \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|_2^2, \left\|\left(f^{(1)}_{\beta^0,j}f_{\beta^0}W\right)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|_1^2\right\}, \end{aligned}$$

and

$$V_{3,j} \le \left[\int_0^\tau \eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)dt\right]^2\times\min\left\{\parallel f_\varepsilon\parallel_\infty\left\|\left(f^2_{\beta^0}W\right)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|_2^2, \left\|\left(f^2_{\beta^0}W\right)^*\frac{K^*_{C_n}}{\overline{f^*_\varepsilon}}\right\|_1^2\right\}.$$

Now, according to Lemma 8.1 we have

$$V_{2,j} \leq \int \left| \left\langle \varphi_2^2(x,\cdot) f_{X,Z}(x,\cdot) \star f_\varepsilon, \left(\left(f^{(1)}_{\beta^0,j} W \right) \star K_{n,C_n} \right)^2 \right\rangle \right| dx$$

and

$$V_{4,j} \leq \int \left| \left\langle \varphi_{4,j}^2(x,\cdot) f_{X,Z}(x,\cdot) \star f_\varepsilon, \left(\left(f_{\beta^0} W \right) \star K_{n,C_n} \right)^2 \right\rangle \right| dx,$$

where

$$\varphi_2(X,Z) = \int_0^\tau \eta_{\gamma^0}(t) dN(t) \text{ and } \varphi_{4,j}(X,Z) = \int_0^\tau \eta^{(1)}_{\gamma^0,j}(t) dN(t).$$

By applying the inequalities (7.1) and (7.2) we get that

$$\begin{aligned} V_{2,j} \leq\ & \mathbb{E}\left[\int_0^\tau \eta_{\gamma^0}(t) dN(t) \right]^2 \\ & \times \min\left\{ \| f_\varepsilon \|_\infty \left\| \left(f^{(1)}_{\beta^0,j} W \right)^* \frac{K^*_{C_n}}{\overline{f^*_\varepsilon}} \right\|_2^2, \left\| \left(f^{(1)}_{\beta^0,j} W \right)^* \frac{K^*_{C_n}}{\overline{f^*_\varepsilon}} \right\|_1^2 \right\}, \end{aligned}$$

and

$$V_{4,j} \leq \mathbb{E}\left[\int_0^\tau \eta^{(1)}_{\gamma^0,j}(t) dN(t) \right]^2 \times \min\left\{ \| f_\varepsilon \|_\infty \left\| \left(f_{\beta^0} W \right)^* \frac{K^*_{C_n}}{\overline{f^*_\varepsilon}} \right\|_2^2, \left\| \left(f_{\beta^0} W \right)^* \frac{K^*_{C_n}}{\overline{f^*_\varepsilon}} \right\|_1^2 \right\}.$$

The result follows by combining the bounds on the $V_{k,j}$'s for $k = 1, \ldots, 4$ and by applying Lemma 8.2 to get that

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial \gamma_j} \right) = O\left(C_n^{2(\alpha-a)+1-\rho+(1-\rho)_-} \exp(-2dC_n^r + 2\delta C_n^\rho))/n \right).$$

The proof of **3)** follows by choosing C_n, that realizes the trade-off between the squared bias and the variance.

Proof of ii)

According to (3.2), $S^{(2)}_{n,1}(\theta^0)$ equals

$$\frac{\partial^2 S_{n,1}(\theta^0)}{\partial \theta^2} = \begin{pmatrix} (S^{(2)}_{n,1})_{1,1} & (S^{(2)}_{n,1})_{1,2} \\ (S^{(2)}_{n,1})^\top_{1,2} & (S^{(2)}_{n,1})_{2,2} \end{pmatrix},$$

with

$$\begin{aligned} (S^{(2)}_{n,1})_{1,2} =\ & -\frac{2}{n} \sum_{i=1}^n (f^{(1)}_{\beta^0} W) \star K_{n,C_n}(U_i) \int_0^\tau (\eta^{(1)}_{\gamma^0}(t))^\top dN_i(t) \\ & + \frac{1}{n} \sum_{i=1}^n \left(f^{(1)}_{\beta^0} f_{\beta^0} W \right) \star K_{n,C_n}(U_i) \int_0^\tau (\eta^{(1)}_{\gamma^0}(t))^\top \eta_{\gamma^0}(t) Y_i(t) dt, \end{aligned}$$

$$\begin{aligned} (S^{(2)}_{n,1})_{1,1} =\ & -\frac{2}{n} \sum_{i=1}^n (f^{(2)}_{\beta^0} W) \star K_{n,C_n}(U_i) \int_0^\tau \eta_{\gamma^0}(t) dN_i(t) \\ & + \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2 (f^2_\beta W)}{\partial \beta^2} \Big|_{\theta=\theta^0} \right) \star K_{n,C_n}(U_i) \int_0^\tau \eta^2_{\gamma^0}(t) Y_i(t) dt \end{aligned}$$

and

$$\begin{aligned}(S^{(2)}_{n,1})_{2,2}(\theta) &= -\frac{2}{n}\sum_{i=1}^n (f_{\beta^0}W)\star K_{n,C_n}(U_i)\int_0^\tau \eta^{(2)}_{\gamma^0}(t)dN_i(t)\\ &\quad+\frac{1}{n}\sum_{i=1}^n (f^2_{\beta^0}W)\star K_{n,C_n}(U_i)\int_0^\tau\left(\frac{\partial^2\eta^2_\gamma(t)}{\partial\gamma^2}\,|_{\theta=\theta^0}\right)Y_i(t)dt.\end{aligned}$$

Under $(\mathbf{A_{15}})$, for C_n satisfying (3.6), $\mathbb{E}[S^{(2)}_{n,1}(\theta^0)-S^{(2)}_{\theta^0,g}(\theta^0)]^2=o(1)$. Hence **ii)** is proved.

Proof of iii)

The proof of **iii)** follows by using the smoothness of $\beta\mapsto Wf_\beta$ and $\beta\mapsto Wf^2_\beta$ up to order 3, the smoothness of $\gamma\mapsto\eta_\gamma$ and $\gamma\mapsto\eta^2_\gamma$ and by using the consistency of $\widehat{\theta}_1$.

Proof of iv)

Let us introduce the random event $E_n=\cap_{j,k}E_{n,j,k}$, where

$$E_{n,j,k}=\left\{\omega \text{ such that } \left|\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}-\frac{\partial^2 S_{n,1}(\theta,\omega)}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}+(R_n)_{j,k}(\omega)\right|\le\frac{1}{2}\,\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}\right\}.$$

We first write that

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}_1-\theta^0\|^2_{\ell^2} &= \mathbb{E}[\|\widehat{\theta}_1-\theta^0\|^2_{\ell^2}\mathbb{1}_{E_n}]+\mathbb{E}[\|\widehat{\theta}_1-\theta^0\|^2_{\ell^2}\mathbb{1}_{E_n^c}]\\ &\le \mathbb{E}[\|\widehat{\theta}_1-\theta^0\|^2_{\ell^2}\mathbb{1}_{E_n}]+2\sup_{\theta\in\Theta}\|\theta\|^2_{\ell^2}\mathbb{P}(E_n^c).\end{aligned}$$

According to (7.10) and (7.11) we have

$$\begin{aligned}\mathbb{E}[\|\widehat{\theta}_1-\theta^0\|^2_{\ell^2}\mathbb{1}_{E_n}] &\le \mathbb{E}\left[(S^{(1)}_{n,1}(\theta^0))^\top[(S^{(2)}_{n,1}(\theta^0)+R_n)^{-1}]^\top(S^{(2)}_{n,1}(\theta^0)+R_n)^{-1}S^{(1)}_{n,1}(\theta^0)\mathbb{1}_{E_n}\right]\\ &\le C(m,p)\sup_{j,k}\left|\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}\right|^{-2}\mathbb{E}\left[(S^{(1)}_{n,1}(\theta^0))^\top S^{(1)}_{n,1}(\theta^0)\right]\\ &\le C(m,p)\sup_{j,k}\left|\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}\right|^{-2}\varphi_n^2\end{aligned}$$

It remains thus to show that $\mathbb{P}(E_n^c)=o(\varphi_n^2)$ with

$$\sup_{j,k}\mathbb{E}\left[\left(\frac{\partial^2(S_{n,1}(\theta)-S_{\theta^0,g}(\theta))}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}\right)^2\right]\le\varphi_n^2.$$

We write

$$\mathbb{P}(E_n^c)\ \le\ \sum_{j=1}^{m+p}\sum_{k=1}^{m+p}\mathbb{P}(E^c_{n,j,k}).$$

By Markov's inequality, for $q>2$,

$$\mathbb{P}(E^c_{n,j,k})\ \le\ \left(\left|\frac{1}{2}\,\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}\right|^q\right)^{-1}\mathbb{E}\left[\left|\left(\frac{\partial^2(S_{\theta^0,g}(\theta)-S_{n,1}(\theta))}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}\right)+(R_n)_{j,k}\right|^q\right].$$

In other words, using that $|a+b|^q\le 2^{q-1}(|a|^q+|b|^q)$, we get

$$\left(\left|\frac{1}{2}\,\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\,|_{\theta=\theta^0}\right|^q\right)\mathbb{P}(E^c_{n,j,k})$$

is less than

$$
\begin{aligned}
& 2^{q-1}\left|\left(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)-\mathbb{E}\left[\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)\right]\right|^q \\
& +2^{q-1}\mathbb{E}\left[\left|\mathbb{E}\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)-\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)+(R_n)_{j,k}\right|^q\right] \\
\leq\ & 2^{q-1}\left|\left(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)-\mathbb{E}\left[\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)\right]\right|^q \\
& +2^{2q-2}\left\{\mathbb{E}\left[\left|\mathbb{E}\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)-\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)\right|^q\right]+\mathbb{E}|(R_n)_{j,k}|^q\right\}.
\end{aligned}
$$

Now we apply the Rosenthal's inequality (see (8.2) recalled in Appendix), to the sum of centered variables

$$
\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)-\mathbb{E}\left[\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)\right] \quad := \quad n^{-1}\sum_{i=1}^n W_{n,i,j,k}.
$$

It follows that

$$
\begin{aligned}
&\mathbb{E}\left[\left|\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)-\mathbb{E}\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)\right|^q\right] \\
&\qquad\qquad \leq C(r)\left[n^{1-r}\mathbb{E}|W_{n,1,j,k}|^q+n^{-q/2}\mathbb{E}^{q/2}|W_{n,1,j,k}|^2\right].
\end{aligned}
$$

Take $q=4$ to get that

$$
\mathbb{E}\left[\left|\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)-\mathbb{E}\left(\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\ |_{\theta=\theta^0}\right)\right)\right|^4\right]\leq C(4)\left[n^{-3}\mathbb{E}|W_{n,1,j,k}|^4+n^{-2}\mathbb{E}^2|W_{n,1,j,k}|^2\right].
$$

Therefore under the conditions ensuring that

$$
\mathbb{E}\Big[\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\theta_k}|_{\theta=\theta^0}-\frac{\partial^2 S_{n,1}(\theta)}{\theta_j\theta_k}|_{\theta=\theta^0}\Big]^2=o(1),
$$

we have

$$
\mathbb{E}\Big[\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\theta_k}|_{\theta=\theta^0}-\frac{\partial^2 S_{n,1}(\theta)}{\theta_j\theta_k}|_{\theta=\theta^0}\Big]^4=O(\varphi_n^4)=o(\varphi_n^2).
$$

Now, by using the definition of R_n and the smoothness properties of the derivatives of (Wf_β) and (Wf_β^2) up to order 3, we get that $\mathbb{E}((R_n)_{j,k}^4)=o(\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^4)$, and we conclude that

$$
\begin{aligned}
\mathbb{E}\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^2 \quad \leq \quad & 4\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top\Big[\Big(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\theta_k}|_{\theta=\theta^0}\Big)^{-1}\Big]^\top\Big(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\theta_k}|_{\theta=\theta^0}\Big)^{-1}S_{n,1}^{(1)}(\theta^0)\right] \\
& +o(\varphi_n^2)+o(\mathbb{E}[\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^4]). \qquad \square
\end{aligned}
$$

7.2. **Proof of Theorem 3.2 : asymptotic normality.** According to Theorem 3.1 and its proof, under $(\mathbf{C_1})$-$(\mathbf{C_3})$, $V_{n,j}(\theta^0)=O(1)$ and the asymptotic normality of $\widehat{\theta}_1$ follows by checking that

v) $\sqrt{n}\ S_{n,1}^{(1)}(\theta^0)\ \xrightarrow[n\to\infty]{\mathcal{L}}\ \mathcal{N}(0,\Sigma_1)$, with Σ_1 defined in Theorem 3.2.

Let $H_{n,i}$, $\widehat{H}_{n,i}$, $G_{n,i}$, and $\widehat{G}_{n,i}$ be the processes defined for all $t\in[0,\tau]$ by

$$
\widehat{H}_{n,i}(s)=\begin{pmatrix}\frac{-2}{\sqrt{n}}(f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(s)\\ \frac{-2}{\sqrt{n}}-(f_{\beta^0}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}^{(1)}(s)\end{pmatrix}, \tag{7.19}
$$

$$H_{n,i}(s)=\begin{pmatrix}\frac{-2}{\sqrt{n}}(f^{(1)}_{\beta^0}W)(Z_i)\eta_{\gamma^0}(s)\\ \frac{-2}{\sqrt{n}}(f_{\beta^0}W)(Z_i)\eta^{(1)}_{\gamma^0}(s)\end{pmatrix}, \tag{7.20}$$

$$\widehat{G}_{n,i}(s)=\begin{pmatrix}\frac{2}{\sqrt{n}}(f_{\beta^0}f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U_i)\eta^2_{\gamma^0}(s)\\ \frac{2}{\sqrt{n}}(f^2_{\beta^0}W)\star K_{n,C_n}(U_i)\eta^{(1)}_{\gamma^0}(s)\eta_{\gamma^0}(s)\end{pmatrix}, \tag{7.21}$$

$$\text{and}\qquad G_{n,i}(s)=\begin{pmatrix}\frac{2}{\sqrt{n}}(f_{\beta^0}f^{(1)}_{\beta^0}W)(Z_i)\eta^2_{\gamma^0}(s)\\ \frac{2}{\sqrt{n}}(f^2_{\beta^0}W)(Z_i)\eta^{(1)}_{\gamma^0}(s)\eta_{\gamma^0}(s)\end{pmatrix}. \tag{7.22}$$

According to (7.12), since $N_i(s)=M_i(s)+\Lambda_i(s,\theta^0,Z_i)$ (see (2.2)), we get that

$$\begin{aligned}\sqrt{n}\,S^{(1)}_{n,1}(\theta^0) &= \sum_{i=1}^n\int_0^\tau \widehat{H}_{n,i}(s)dN_i(s)+\sum_{i=1}^n\int_0^\tau \widehat{G}_{n,i}(s)Y_i(s)ds\\ &= A_1+A_2+A_3+A_4\end{aligned}$$

with

$$\begin{aligned}A_1 &= \sum_{i=1}^n\int_0^\tau H_{n,i}(s)dM_i(s),\quad A_2=\sum_{i=1}^n\int_0^\tau[\widehat{H}_{n,i}(s)-H_{n,i}(s)]dM_i(s),\\ A_3 &= \sum_{i=1}^n\int_0^\tau[\widehat{H}_{n,i}(s)-H_{n,i}(s)]d\Lambda_i(s,\theta^0,Z_i) \text{ and } A_4=\sum_{i=1}^n\int_0^\tau[\widehat{G}_{n,i}(s)-G_{n,i}(s)]Y_i(s)ds.\end{aligned}$$

Study of A_1

The term A_1 is a linear combinations of stochastic integrals of locally bounded and predictable processes, $H_{n,i}$, with respect to finite variation and local square integrable martingales, $M_i(\cdot)$. Consequently, $\mathbb{E}(A_1)=0$. Denoting by $<M>$ the predictable variation process of M we have to verify the two following conditions for all t in $[0,\tau]$ (see (Andersen, Borgan, Gill, and Keiding 1993) page 68) :

L1) $\sum_{i=1}^n\int_0^t H_{n,i}(s)(H_{n,i}(s))^\top d<M_i>(s)\ \xrightarrow[n\to\infty]{\mathbb{P}}\ \widetilde{\Sigma}^2_1(t)$, with $\widetilde{\Sigma}^2_1(t)$ a positive covariance matrix defined by

$$\tilde{\Sigma}^2_1(t)=4\mathbb{E}\left[\int_0^t\begin{pmatrix}(f^{(1)}_{\beta^0}W)(Z_i)\eta_{\gamma^0}(s)\\(f_{\beta^0}W)(Z_i)\eta^{(1)}_{\gamma^0}(s)\end{pmatrix}\begin{pmatrix}(f^{(1)}_{\beta^0}W)(Z_i)\eta_{\gamma^0}(s)\\(f_{\beta^0}W)(Z_i)\eta^{(1)}_{\gamma^0}(s)\end{pmatrix}^\top\eta_{\gamma^0}(s)Y_i(s)ds\right] \tag{7.23}$$

L2) For all $\epsilon>0$, $\sum_{i=1}^n\int_0^t H_{n,i}(s)(H_{n,i}(s))^\top\mathbb{1}_{\|H_{n,i}(s)\|_{\ell^2}\geq\epsilon}\ d<M_i>(s)=o_p(1)$.

Proof of **L1)**

Since $<M_i>=\Lambda_i$, we have to prove that for all $t\in[0,\tau]$,

$$\sum_{i=1}^n\int_0^t H_{n,i}(s)(H_{n,i}(s))^\top Y_i(s)f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)ds\ \xrightarrow[n\to\infty]{\mathbb{P}}\ \widetilde{\Sigma}^2_1(t). \tag{7.24}$$

We apply the following Lemma, which is a straightforward consequence of the fact that the set of functions $\mathcal{I}_t=\{x\mapsto\mathbb{1}_{x\geq t}\}$ is a $\mathbb{P}$-Glivenko Cantelli class (see van der Vaart and Wellner (1996)).

Lemma 7.1. *For $j=1,\cdots,m$*

$$\sup_{0\leq t\leq\tau}\left|\frac{1}{n}\sum_{i=1}^n Y_i(t)f_{\beta^0,j}(Z_i)(f^{(1)}_{\beta^0,j}W)(Z_i)-\mathbb{E}[Y(t)f_{\beta^0}(Z)(f^{(1)}_{\beta^0,j}W)(Z)]\right|\ \xrightarrow[n\to\infty]{P.S.}\ 0,$$

$$\sup_{0\le t\le\tau}\left|\frac{1}{n}\sum_{i=1}^{n}Y_i(t)f^2_{\beta^0}(Z_i)W(Z_i)-\mathbb{E}[Y(t)f^2_{\beta^0}(Z)W(Z)]\right|\xrightarrow[n\to\infty]{P.S.}0$$

$$\sup_{0\le t\le\tau}\left|\frac{1}{n}\sum_{i=1}^{n}Y_i(t)f_{\beta^0}(Z_i)|(f^{(1)}_{\beta^0,j}W)(Z_i)|^3-\mathbb{E}[Y(t)f_{\beta^0}(Z)|(f^{(1)}_{\beta^0,j}W)(Z)|^3]\right|\xrightarrow[n\to\infty]{P.S.}0,$$

$$\textit{and}\ \sup_{0\le t\le\tau}\left|\frac{1}{n}\sum_{i=1}^{n}Y_i(t)|f_{\beta^0}(Z_i)W(Z_i)|^3-\mathbb{E}[Y(t)|f_{\beta^0}(Z)W(Z)|^3]\right|\xrightarrow[n\to\infty]{P.S.}0.$$

Thus **L1)** is checked .
Proof of **L2)**. We have to check that for all $j=1,\dots,m$

$$\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\int_0^t\left[f^{(1)}_{\beta^0,j}(Z_i)W(Z_i)\eta_{\gamma^0}(s)\right]^2\mathbb{1}_{|f^{(1)}_{\beta^0,j}(Z_i)W(Z_i)\eta_{\gamma^0}(s)|\ge\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]=o(1)$$

and that for all $j=1,\dots,p$

$$\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\int_0^t\left[f^2_{\beta^0}(Z_i)W(Z_i)\eta^{(1)}_{\gamma^0,j}(s)\right]^2\mathbb{1}_{|f^2_{\beta^0}(Z_i)W(Z_i)\eta^{(1)}_{\gamma^0,j}(s)|\ge\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]=o(1).$$

This is a straightforward consequence of Lemma 7.2 by writing that for $j=1,\cdots,m$

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\int_0^t\left[f^{(1)}_{\beta^0,j}(Z_i)W(Z_i)\eta_{\gamma^0}(s)\right]^2\mathbb{1}_{|f^{(1)}_{\beta^0,j}(Z_i)W(Z_i)\eta_{\gamma^0}(s)|\ge\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]\\&\qquad\le\frac{1}{n\sqrt{n}\epsilon}\mathbb{E}\left[\sum_{i=1}^{n}\int_0^t|f^{(1)}_{\beta^0,j}(Z_i)W(Z_i)\eta_{\gamma^0}(s)|^3f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]=o(1)\end{aligned}$$

and for $j=1,\cdots,p$

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\int_0^t\left[f^2_{\beta^0}(Z_i)W(Z_i)(\eta^{(1)}_{\gamma^0,j}(s)\right]^2\mathbb{1}_{|f^2_{\beta^0}(Z_i)W(Z_i)\eta^{(1)}_{\gamma^0,j}(s)|\ge\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]\\&\qquad\le\frac{1}{\epsilon n\sqrt{n}}\mathbb{E}\left[\sum_{i=1}^{n}\int_0^t|f^2_{\beta^0}(Z_i)W(Z_i)|^3|\eta^{(1)}_{\gamma^0,j}(s)|^3f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]=o(1).\end{aligned}$$

Thus **L2)** is checked.

Study of A_2

Since $\mathbb{E}(A_2)=0$, we use the following lemma, analogous to Lemma 7.1.

Lemma 7.2. *Under* **(A$_{14}$)**-**(A$_{15}$)**, *for C_n satisfying (3.6) thenfor $j=1,\cdots,m$*

$$\sup_{0\le t\le\tau}\left|\frac{1}{n}\sum_{i=1}^{n}Y_i(t)f_{\beta^0}(Z_i)(f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_i)-\mathbb{E}[Y(t)f_{\beta^0}(Z)(f^{(1)}_{\beta^0,j}W)(Z)]\right|\xrightarrow[n\to\infty]{P.S.}0,$$

$$\textit{and}\ \sup_{0\le t\le\tau}\left|\frac{1}{n}\sum_{i=1}^{n}Y_i(t)f_{\beta^0}(Z_i)(f_{\beta^0}W)\star K_{n,C_n}(U_i)-\mathbb{E}[Y(t)f^2_{\beta^0}(Z)W(Z)]\right|\xrightarrow[n\to\infty]{P.S.}0.$$

It follows that $A_2=o_p(1)$.

Study of A_3

It is noteworthy that the term A_3 can be seen as triangular arrays of row-wise independent centered random variables that is

$$A_3 = \sum_{i=1}^n V_{n,i} + \mathbb{E}(A_3),$$

with $\sum_{i=1}^n V_{n,i} = A_3 - \mathbb{E}(A_3)$. Consequently, the asymptotic normality follows by checking that

v-a) $\mathbb{E}(A_3) = o_p(1)$

v-b) $\sum_{i=1}^n \mathbb{E}[(V_{n,i})^2] \underset{n\to\infty}{\longrightarrow} \Sigma_3^2$

v-c) For all $\epsilon > 0$, $\sum_{i=1}^n \mathbb{E}[(V_{n,i})^2 \mathbb{1}_{\|V_{n,i}\|_{\ell^2}\geq \epsilon}] \underset{n\to\infty}{\longrightarrow} 0$ (Lindeberg Condition).

By definition, A_3 equals

$$-\frac{2}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \begin{pmatrix} (f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U_i) - (f^{(1)}_{\beta^0}W)(Z_i)\eta_{\gamma^0}(s) \\ (f_{\beta^0}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}W)(Z_i)\eta^{(1)}_{\gamma^0}(s)\end{pmatrix} Y_i(s) f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)ds.$$

Let us start with the study of the variance (**v-b**). Under $(\mathbf{C_1})$-$(\mathbf{C_3})$

$$\mathrm{Var}\left[-\frac{2}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau (f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U_i) - (f^{(1)}_{\beta^0}W)(Z_i)\eta_{\gamma^0}(s)Y_i(s)f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)ds\right] = O(1),$$

and

$$\mathrm{Var}\left[-\frac{2}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau (f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}W)(Z_i)\eta^{(1)}_{\gamma^0}(s)Y_i(s)f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)ds\right] = O(1).$$

It follows that **v-b)** is checked.

We now come to the bias term and write that

$$\begin{aligned}\mathbb{E}(A_3) &= \mathbb{E}\left\{-\frac{2}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \begin{pmatrix} (f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U_i) - (f^{(1)}_{\beta^0}W)(Z_i)\eta_{\gamma^0}(s) \\ (f_{\beta^0}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}W)(Z_i)\eta^{(1)}_{\gamma^0}(s)\end{pmatrix} Y_i(s)f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)\right\} ds \\ &= -2\sqrt{n}\begin{pmatrix} \mathbb{E}\left\{\left[(f^{(1)}_{\beta^0}W)\star K_{n,C_n}(U) - (f^{(1)}_{\beta^0}W)(Z)\right] f_{\beta^0}(Z_i)\int_0^\tau \eta^2_{\gamma^0}(s)Y(s)ds\right\} \\ \mathbb{E}\left\{\left[(f_{\beta^0}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}W)(Z_i)\right] f_{\beta^0}(Z_i)\int_0^\tau \eta^{(1)}_{\gamma^0}(s)\eta_{\gamma^0}(s)Y(s)ds\right\}\end{pmatrix}.\end{aligned}$$

According to Lemma 8.1

$$\begin{aligned}\mathbb{E}(A_3) &= -2\sqrt{n}\begin{pmatrix} \mathbb{E}\left\{\left[(f^{(1)}_{\beta^0}W)\star K_{C_n}(Z) - (f^{(1)}_{\beta^0}W)(Z)\right] f_{\beta^0}(Z)\int_0^\tau Y(s)\eta^2_{\gamma^0}(s)ds\right\} \\ \\ \mathbb{E}\left\{\left[(f_{\beta^0}W)\star K_{C_n}(Z) - (f_{\beta^0}W)(Z)\right] f_{\beta^0}(Z)\int_0^\tau Y(s)\eta^{(1)}_{\gamma^0}(s)\eta_{\gamma^0}(s)ds\right\}\end{pmatrix} \\ &= -2\sqrt{n}\begin{pmatrix} \mathbb{E}\left\{\left[(f^{(1)}_{\beta^0}W)\star K_{C_n}(Z) - (f^{(1)}_{\beta^0}W)(Z)\right] f_{\beta^0}(Z)\int_0^\tau Y(s)\eta^2_{\gamma^0}(s)ds\right\} \\ \\ \mathbb{E}\left\{\left[(f_{\beta^0}W)\star K_{C_n}(Z) - (f_{\beta^0}W)(Z)\right] f_{\beta^0}(Z)\int_0^\tau Y(s)\eta^{(1)}_{\gamma^0}(s)\eta_{\gamma^0}(s)ds\right\}\end{pmatrix} \\ &= -2\sqrt{n}\begin{pmatrix} \int\left\langle (f^{(1)}_{\beta^0}W)\star K_{C_n}(z) - (f^{(1)}_{\beta^0}W)(z), f_{\beta^0}(z)f_{X,Z}(x,z)\right\rangle \left(\int_0^\tau \mathbb{1}_{x\geq s}\eta^2_{\gamma^0}(s)ds\right) dx \\ \\ \int\left\langle (f_{\beta^0}W)\star K_{C_n}(z) - (f_{\beta^0}W)(z), f_{\beta^0}(z)\right\rangle \left(\int_0^\tau \mathbb{1}_{x\geq s}\eta^{(1)}_{\gamma^0}(s)\eta_{\gamma^0}(s)ds\right) dx\end{pmatrix}.\end{aligned}$$

For $j = 1,\ldots,m$

$$\left|\int \left\langle (f^{(1)}_{\beta^0,j}W)\star K_{C_n}(z) - (f^{(1)}_{\beta^0,j}W)(z), f_{\beta^0}(z)f_{X,Z}(x,z)\right\rangle \left(\int_0^\tau \mathbb{1}_{x\geq s}\eta^2_{\gamma^0}(s)ds\right) dx\right|$$

is less than

$$\left(\int_0^\tau \eta^2_{\gamma^0}(s)ds\right)\int\left|\left\langle (f^{(1)}_{\beta^0,j}W)\star K_{C_n}(z)-(f^{(1)}_{\beta^0,j}W)(z), f_{\beta^0}(z)f_{X,Z}(x,z)\right\rangle\right|dx,$$

which is, according to (7.1) and (7.2), less than

$$\left(\int_0^\tau \eta^2_{\gamma^0}(s)ds\right)\min\left\{\int\left\|(f^{(1)}_{\beta^0,j}W)\star K_{C_n}-(f^{(1)}_{\beta^0,j}W)\right\|_2\left\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\right\|_2dx,\right.$$
$$\left.\int\left\|(f^{(1)}_{\beta^0,j}W)\star K_{C_n}-(f^{(1)}_{\beta^0,j}W)\right\|_\infty\left\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\right\|_1dx\right\}$$

that is less than

$$(2\pi)^{-1}\left(\int_0^\tau \eta^2_{\gamma^0}(s)ds\right)\times\min\left\{\left\|(f^{(1)}_{\beta^0,j}W)^*(K^*_{C_n}-1)\right\|_2\int\left\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\right\|_2dx,\right.$$
$$\left.\left\|(f^{(1)}_{\beta^0,j}W)^*(K^*_{C_n}-1)\right\|_\infty\int\left\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\right\|_1dx\right\}.$$

In the same way we obtain that for $j=1,\ldots,p$

$$\left|\int\left\langle (f_{\beta^0}W)\star K_{C_n}(z)-(f_{\beta^0}W)(z), f_{\beta^0}(z)\right\rangle\left(\int_0^\tau \mathbb{1}_{x\geq s}\eta^{(1)}_{\gamma^0}(s)\eta_{\gamma^0}(s)ds\right)dx\right|$$

is less than

$$(2\pi)^{-1}\left(\int_0^\tau |\eta_{\gamma^0}(s)\eta^{(1)}_{\gamma^0,j}(s)|ds\right)\times\min\left\{\left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_2\int\left\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\right\|_2dx,\right.$$
$$\left.\left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_\infty\int\left\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\right\|_1dx\right\}.$$

Consequently, under $(\mathbf{A_{15}})$, $\mathbb{E}(A_3)=O(\sqrt{n}C_n^{-a+(1-r)/2+(1-r)_-/2\exp(-dC_n^r)})$. Under $(\mathbf{C_1})$-$(\mathbf{C_3})$, $\mathrm{Var}(A_3)=O(1)$ and hence C_n can be chosen such that $\mathbb{E}(A_3)=o(1)$. It follows that **v-a)** is checked.

In order to check the Lindeberg condition we write that for $j=1,\cdots,m$

$$\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n\int_0^t\left[(f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(s)\right]^2\mathbb{1}_{|(f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(s)|\geq\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]$$
$$\leq\frac{1}{n\sqrt{n}\epsilon}\mathbb{E}\left[\sum_{i=1}^n\int_0^t|(f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(s)|^3f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]=o(1)$$

and for $j=1,\cdots,p$

$$\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n\int_0^t\left[(f^2_{\beta^0}W)\star K_{n,C_n}(U_i)\eta^{(1)}_{\gamma^0,j}(s)\right]^2\mathbb{1}_{|(f^2_{\beta^0}W)\star K_{n,C_n}(U_i)\eta^{(1)}_{\gamma^0,j}(s)|\geq\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]$$
$$\leq\frac{1}{\epsilon n\sqrt{n}}\mathbb{E}\left[\sum_{i=1}^n\int_0^t|(f^2_{\beta^0}W)\star K_{n,C_n}(U_i)|^3|\eta^{(1)}_{\gamma^0,j}(s)|^3f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\right]=o(1).$$

It follows that **v-c)** is checked.

Study of A_4

The study of A_4, quite similar to the study of A_3 is omitted. □

7.3. **Proof of Theorem 4.1 :** The proof of Thorem 4.1, quite classical is omitted.

8. Appendix

Lemma 8.1. *Let φ be such that $\mathbb{E}(|\varphi(X,Z)|)$ is finite and let Φ such that $\mathbb{E}(|\Phi(U)|)$ is finite. Under the assumptions ($\mathbf{A_3}$) and ($\mathbf{A_4}$), then*

$$\mathbb{E}[\varphi(X,Z)\Phi\star K_{n,C_n}(U)] = \mathbb{E}[\varphi(X,Z)\Phi\star K_{C_n}(Z)],$$

and

$$\mathbb{E}\left[\varphi(X,Z)\Phi\star K_{n,C_n}(U)\right]^2 = \int \left\langle \left(\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)\right)\star f_\varepsilon, (\Phi\star K_{n,C_n})^2\right\rangle dx.$$

Proof of Lemma 8.1 : Set $f_{X,U,Z}$ the joint distribution of (X,U,Z). Under ($\mathbf{A_3}$) and ($\mathbf{A_4}$), $f_{X,U,Z}(x,u,z) = f_{X,Z}(x,z)f_\varepsilon(u-z)$. Hence by the Parseval's formula

$$\begin{aligned}
\mathbb{E}\left[\varphi(X,Z)\Phi\star K_{n,C_n}(U)\right] &= \iiint \varphi(x,z)\Phi\star K_{n,C_n}(u)f_{X,Z}(x,z)f_\varepsilon(u-z)du\,dx\,dz\\
&= \iint \varphi(x,z)f_{X,Z}(x,z)\int \Phi\star K_{n,C_n}(u)f_\varepsilon(u-z)du\,dx\,dz\\
&= (2\pi)^{-1}\iint \varphi(x,z)f_{X,Z}(x,z)\int \Phi^*(y)K^*_{n,C_n}(y)f^*_\varepsilon(y)e^{-iyz}dy\,dx\,dz\\
&= (2\pi)^{-1}\iint \varphi(x,z)f_{X,Z}(x,z)\int \Phi^*(y)\frac{K^*_{C_n}(y)}{\overline{f^*_\varepsilon}(y)}\overline{f^*_\varepsilon}(y)e^{-iyz}dy\,dx\,dz\\
&= (2\pi)^{-1}\iint \varphi(x,z)f_{X,Z}(x,z)\int \Phi^*(y)K^*_{C_n}(y)e^{-iyz}dy\,dx\,dz\\
&= \iint \varphi(x,z)f_{X,Z}(x,z)\int \Phi(u)K_{C_n}(z-u)du\,dx\,dz\\
&= \iint \varphi(x,z)\Phi\star K_{C_n}(z)f_{X,Z}(x,z)\,dx\,dz.
\end{aligned}$$

In the same way,

$$\begin{aligned}
\mathbb{E}\left[\varphi(X,Z)\Phi\star K_{n,C_n}(U)\right]^2 &= \iiint \varphi^2(x,z)(\Phi\star K_{n,C_n}(u))^2 f_{X,Z}(x,z)f_\varepsilon(u-z)dx\,du\,dz\\
&= \iiint \varphi^2(x,z)(\Phi\star K_{n,C_n}(u))^2 f_{X,Z}(x,z)f_\varepsilon(u-z)dx\,du\,dz\\
&= \int \left\langle (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot))\star f_\varepsilon, (\Phi\star K_{n,C_n})^2\right\rangle dx. \qquad \square
\end{aligned}$$

Lemma 8.2. *For a, r two nonnegative numbers, Then*

$$\int_{|u|\geq C_n} |u|^{-\nu}\exp(-\lambda|u|^\delta)du \leq \frac{1}{C(\nu,\lambda,\delta)}C_n^{-\nu+1-\delta}\exp\{-\lambda C_n^\delta\}. \tag{8.1}$$

Furthermore, if f_ε satisfies ($\mathbf{A_{14}}$), then

$$\int_{|u|\leq C_n}\frac{|u|^{-\nu}\exp(-\lambda|u|^\delta)}{|f^*_\varepsilon(u)|}du \leq \frac{1}{C(\alpha,\delta,\rho,\nu,\lambda,\delta)\underline{C}(f_\varepsilon)}\max[1, C_n^{(\alpha-\nu+1-\delta)}\exp\{-\lambda C_n^\delta+\delta C_n^\rho\}].$$

Lemma 8.3. ***Rosenthal's inequality*** *(Rosenthal (1970), Petrov (1995)). For $U_1,\ldots,U_n$, be n independent centered random variables, there exists a constant $C(r)$ such that for $r\geq 1$,*

$$\mathbb{E}[|\sum_{i=1}^n U_i|^r] \leq C(r)[\sum_{i=1}^n \mathbb{E}[|U_i|^r] + (\sum_{i=1}^n \mathbb{E}[U_i^2])^{r/2}]. \tag{8.2}$$

References

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical models based on Counting processes.* Springer Series in Statistics.

Augustin, T. (2004). An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scand. J. Stat. 31*(1), 43–50.

Buzas, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *J. Statist. Plann. Inference 67*, 247–257.

Carroll, R., Ruppert, D., and Stefanski, L. (1995). *Measurement error in nonlinear models.* Chapman and Hall, London.

Fedoryuk, M. V. (1987). *Asimptotika: integraly i ryady.* Spravochnaya Matematicheskaya Biblioteka. [Mathematical Reference Library]. Moscow: "Nauka".

Fuller, W. A. (1987). *Measurement error models.* Wiley Series in Probability and Mathematical Statistics.

Gill, R. D. and Andersen, P. K. (1982). Cox's regression model for counting processes : a large sample study. *Ann. Statist. 10*, 1100–1120.

Gong, G., Whittemore, A. S., and Grosser, S. (1990). Censored survival data with misclassified covariates: A case study of breast-cancer mortality. *J. Amer. Statist. Assoc. 85*(409), 20–28.

Hu, C. and Lin, D. (2002). Cox regression with covariate measurement error. *Scand. J. Stat. 29*(4), 637–655.

Huang, Y. and Wang, C. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *J. Am. Stat. Assoc. 95*(452), 1209–1219.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Annals math. Statist. 27*, 887–906.

Kong, F. H. (1999). Adjusting regression attenuation in the cox proportional hazards model. *J. Statist. Plann. Inference 79*, 31–44.

Kong, F. H. and Gu, M. (1999). Consistent estimation in cox proportional hazards model with covariate measurement errors. *Statistica Sinica 9*, 953–969.

Lepski, O. V. and Levit, B. Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist. 7*(2), 123–156.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrika 77*, 127–137.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics 48*, 829–838.

Pons, O. (2002). Estimation in the Cox model with missing covariate data. *J. Nonparametric Stat. 14*(3), 223–247.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 331–342.

Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for cox-type regression models with general relative risk form. *Ann. Statist. 11*, 804–813.

Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica. 18*, 375–389.

Stefanski, L. A. (1989). Unbiaised estimation of a nonlinear function of a normal mean with application to measurement error models. *Commun. Statist. -Theory Meth. 18*, 4335–4358.

Tsiatis, T. T., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. application to survival and cd4 counts in patients with aids. *J. Amer. Statist. Assoc. 90*, 27–37.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergences and empirical processes. With applications to Statistics*. Springer Series in Statistics, New York.

Marie-Laure MARTIN-MAGNIETTE
Institut National Agronomique Paris-Grignon
Mathématique et Informatique Appliquées,
16, rue Claude Bernard
75231 Paris cedex 05, France,

Institut National de la Recherche Agronomique,
Unité de Recherche en Génomique Végétale,
UMR INRA 1165- CNRS 8114 -UEVE
2,rue Gaston Crémieux- CP 5708
91057 Evry Cedex, France
e-mail : mlmartin@inapg.fr

Marie-Luce TAUPIN
Laboratoire de Probabilités, Statistique et Modélisation, UMR 8628,
Université Paris-Sud, Bât. 425,
91405 Orsay Cedex, France
e-mail : marie-luce.taupin@math.u-psud.fr